Design and Development of a Knowledge Modelling Approach to Govern the Use of Electronic Health Records for Research

Nathan C. Lea

Thesis submitted in accordance with the requirements of the University of London for the degree of Doctor of Philosophy

University College London

2015

Declaration:

I, Nathan Christopher Lea confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Dedication

This thesis is dedicated to Rene Branton:
neighbour, scholar, mentor and friend.

Abstract

There is now increasing commitment internationally to using electronic healthcare records collected during routine care delivery to conduct clinical research. This must be rigorously controlled by an extensive set of information governance requirements defining the legal, ethical and practical guidelines to respect the privacy rights of the people about whom the records are kept, uphold the clinical profession's duty of confidentiality and protect the interests of participants, practitioners and researchers. The development of information security policies is a highly regarded method of meeting these requirements. This is hampered by the need to interpret a complex framework of legislation and guidelines, lack of clear advice and inconsistency in authoring, interpretation and understanding amongst the people whose behaviour they are expected to guide.

By using the results of several UK and European research and information platform development projects in which the author has participated and by gathering requirements from stakeholders in the clinical and research communities, this thesis defines a knowledge management representation to specify policy requirements in a computable form. The work provides the first set of knowledge requirements for governing research uses of electronic healthcare records, and a knowledge model that describes information security policies and generates a web application tool. The tool allows policy control authoring that provides a consistent, clear and unambiguous view of governance requirements to researchers and service providers.

The model and tool have been evaluated in a laboratory setting to explore their effects on behaviour and understanding of invited participants when authoring policy about handling healthcare records in research and making decisions about sharing information. The work has resulted in a validation of the model and demonstrated the potential positive effects of this new approach on practice. It makes recommendations about how it should be used in working practice and for educating people about information governance when performing clinical research to improve care provision.

Table of Contents

List of Figures

List of Tables

List of Graphs

List of Code Scripts

Acknowledgements

This thesis has come to realisation through the supervision, support, guidance and friendship of Professor Dipak Kalra and Professor Stephen Hailes, who have inspired and helped me to develop my career and growth in academia with patience, generosity and trust. The work would not have developed and matured without the support and mentoring of my friend and colleague Dr. Tony Austin, who has enriched this journey with intellect, engineering prowess and good humour. I offer my thanks and appreciation to Professor David Ingram and Professor David Patterson, who have helped and guided me with a keen and supportive interest in the progress of the work and my career, and to Dr. Jackie Nicholls for her mentoring and guidance. I also wish to offer my thanks to the research participants who were very kind in taking time out of their busy schedules to support this research and without whom none of this would be meaningful or possible.

My gratitude to my colleagues in the Centre for Health Informatics and Multiprofessional Education for their help and confidence in me, and particular thanks to the Electronic Healthcare Records Team: Dr. Archana Tapuria, Ms. Yin Su Lim and Mr. David Nguyen. I have been privileged to work with supportive colleagues during the development of this thesis: Mr. Peter Singleton, Professor Donia Scott, Dr. James Cunningham, Dr. Brock Craft, Professor Elizabeth Murray, Ms. Karen Tingay, Dr. Jacky Pallas, Mr. Anthony Peacock, Dr. Spiros Denaxas, Professor Harry Hemingway, Dr. Mark Leaning and Ms. Rae Harbird. Special mention must go to our "PhD Student Support Group," Dr. Bridget Coleman, Dr. Chris Martin, Dr. Cicely Kerr, Dr. Caroline McGraw and Dr. Zarnie Khadjesari.

To my family and friends: I am fortunate and deeply grateful to have so many of you who deserve to be acknowledged here that I could not hope to list you all and keep within the word limit. Each of you have kept me laughing, honest and inspired and I have the upmost faith in our relationships that you know what you have done and how much you mean to me. My acknowledgements would however not be complete without thanking my mother, Heather, who worked tirelessly to support, raise and educate me so that I could achieve what I have with this work.

“Health Informatics is 80% People”

Professor Don Detmer, London, May 2013

Chapter 1. Introduction

This thesis proposes a knowledge management framework for enabling the clinical research community to adhere to safe, legal and ethical practice when using electronic healthcare records to perform their research. The goal of the framework is to simplify the process of developing information security policies that guide people engaged in performing research so that all members of research teams can achieve a consistent, clear and correct understanding of what is required of them. These requirements are based on the legal, ethical and practical expectations for protecting the individuals about whom the information has been recorded when they present for healthcare services. The care they receive is delivered under the clinical profession's duty of confidentiality, which is the basis of a relationship of trust between the patient and their attending clinician. The recorded information is subject to a Common law Duty of Confidentiality, a legally protected right to personal privacy and data protection laws. These legal protections form part of a framework of good working practice expectations and international standards for information security management, which together represent the basis of information governance for these records.

There are compelling reasons to share these records for purposes other than healthcare provision, the goals of which are to improve healthcare services and outcomes for patients. The information that has been collected during care delivery has been for the purposes of care but not these other purposes, which include clinical research. The current method of information governance for these purposes is consequently subject to a plethora of approvals, legal compliance requirements and the development of multiple information governance policies and agreements. These have to be interpreted, understood and put into practice by people who are using healthcare information for performing clinical research as well as refined to configure information security software tools. The literature on this topic and the work in this thesis have identified that this is a complex set of requirements, which have been prone to limited understanding, wide interpretation and uncertainty about best practice. There has been no clear, unified method or resource to guide people on how to develop information

governance controls, express them in a way that offers a consistent view between different users and research teams using the same data whilst helping to achieve a clear understanding of these requirements.

Knowledge management in the area of electronic healthcare information systems development has been a highly anticipated and broadly successful approach to encourage a timely, consistent understanding between a variety of clinical specialists for various uses that support caring for patients (Beale, 2002, Garde et al., 2007a, Kalra, 2002, Kalra and Fernando, 2013, Delaney, 2009). By developing a set of models that meet the requirements for information governance domain knowledge and using this to develop a tool for advising a variety of research users on expected behaviour, the author proposed that a knowledge management framework clarified the legal, ethical and good practice requirements for people when they specified policy and handled healthcare information for clinical research. The work aimed to evaluate the approach as a proof of concept and test the following hypotheses, which proposed that the knowledge management framework:

1. encouraged a consistent understanding of expected behaviour across a range of role holders when authoring and reviewing information governance policy;
2. limited variation in interpreting information governance requirements when authoring and reviewing policies;
3. supported user expertise when interpreting required behaviour and refined these requirements to computable heuristics.

The results of this work have been to define a set of requirements to manage information governance policies and procedures. These requirements have been used to develop and implement a knowledge model that represents the information and structure needed to support effective information governance. The knowledge model has been used to develop a tool that allows users to author a computable and human readable representation of policies according to a consistent structure, which can then be used to advise others on how they should behave with information assets according to an agreed policy. The model and tool

have been evaluated in a laboratory environment using examples of "real world," independently written information security policies, across a range of role holders responsible for running and conducting research using electronic healthcare records. The evaluation of the approach as a proof of concept has involved the use of human participants to test these hypotheses and gather results about the their views on using it and whether they would use it in practice. These evaluations provided evidence of the ease of policy authoring and the quality and consistency of those authored policies when used by the research participants.

This chapter introduces the research work, providing a background to the research area. This includes an introduction to the information governance frameworks in place to protect healthcare information as it is shared and used for other purposes. It also introduces the support for healthcare information sharing that has been increasing over the last twenty years, the problems that have developed and motivations for this research work. It concludes with a summary of the thesis structure.

1.1. Information Governance of Healthcare Records

The use of information gathered from individuals is governed by stringent legal and ethical constraints that represent the clinical profession's duty of confidentiality and the individual's rights to privacy, recognised internationally in human rights and data protection laws, described in section 3.1. These constraints are applied using information security techniques described in section 3.2, which involve the development of procedures and practices that people who are responsible for processing sensitive information and its protection must be made aware of. The management of legal, ethical and good practice requirements using information security techniques is known as information governance in the UK. The United Kingdom is considered as a recurrent case study in this thesis: it has proved to be a leading example of pioneering information sharing trends internationally because the National Health Service (NHS) operates the majority of health services for the UK population of approximately fifty eight million people and collects information under the same legal, organisational and ethical framework.

The legal requirements in the UK and internationally have been identified in the literature as being poorly understood, complex and unclear, resulting in variable interpretations to the detriment of care and research (Academy of Medical Sciences, 2006, Laurie and Sethi, 2013), which have prompted ongoing debate over the identified issues (Times Newspaper, 2006, Bobrow, 2013) and a review of data sharing practice (Thomas and Walport, 2008). Successive governments in the UK have adapted existing and adopted new legislation and guidelines to support greater sharing of information to support healthcare service provision, whilst attempting to balance the fundamental, individual rights to privacy and the duty of confidentiality core to the clinical profession, as discussed in section 3.1.2. These amendments in legislation have caused anxieties about civil liberties and individual rights to privacy amongst the public, prompting a series of reports and petitions to try to limit the amendments and the sharing of information that was proposed, whilst encouraging a greater expectation of accountability to those who curate healthcare records, as discussed in section 3.3.

In response to these anxieties the UK Secretary of State for Health in 2012 commissioned an independent review of information governance (Department of Health, 2013b), which made a series of recommendations for the handling of health and social care information. For secondary uses such as clinical research and commissioning, it recommended that organisations processing and linking de-identified healthcare information should do so with an appropriate legal basis in "accredited safe havens," discussed in section 4.3, where accreditation would require independent and routine audit, compliance and certification with components of standards such as the International Organization for Standardization (ISO) 27000 series of standards on information security (International Organisation for Standardization (ISO), 2014).

The ISO 27000 series has been developed to provide a framework for guiding organisations on how to establish good information security management and a code of practice for implementing it. It recommends that organisations ensure that management is committed to information security, mandates that risk assessments be run and that an Information Security Management System (ISMS) is established using different stakeholders within an organisation to oversee, periodically review,

revise and enact security controls to help mitigate risks to the organisation and maintain legal compliance as information is processed.

The authoritative control mechanism at the disposal of the ISMS is a framework of policies developed to guide organisational members on how to achieve safe working practice as well as handle information assets. The policy framework contains the detail that is needed to inform people of how to behave, as well as defining responsibility, policy review and action that must be taken to prevent a breach or report issues. It is generally accepted that people are the most significant risk to an organisation: any controls must therefore be targeted at the people whose behaviour needs guidance if they are to be successful.

1.2. Information Security Policies

ISO IEC 27001: 2013 defines the requirements for information security management, mandating that policies must be written in a way that is "relevant, accessible and understandable to the intended reader" (British Standards Institute, 2013a). The primary role of an information security policy is to guide people who are using information in how to meet their responsibilities to protect it. Policies provide a reference framework for how people should behave in order to achieve the required protection. They must therefore encompass all aspects that are involved with the management of the information, including details of who may be involved with curating and processing the information, what they are responsible for, how the information is processed and a register of assets that are available. They must reflect changes in working practice and activities as the goals change and evolve, as well as any changes in overarching legislation, organisational policy and published good practice.

When developing an information security plan and policy, a risk assessment should be conducted to identify individuals within the organisation who own those risks, vulnerabilities in the information assets and threats, which leads to the specification of a mitigation strategy using available protection mechanisms. Threats must be comprehensively identified so that a thorough mitigation strategy is developed based upon the likelihood of threat compromising an asset and the estimated impact on reputation, business continuity and / or financially. The

mitigation strategies must cover everything from the specification of computerised access control policy for information management software systems to the physical security of the rooms where the computer hardware that stores the information and runs processing software are placed. The details of a risk assessment and mitigation strategies should inform the development of policies.

The governance of clinical research projects involving healthcare information tends to focus their policy and risk mitigation requirements around maintaining the anonymity of participants. The controls therefore include de-identification techniques described in section 3.1.1, which tend to be applied in levels: higher levels of de-identification, often described as anonymisation, make information virtually anonymous but also potentially less useful for a variety of purposes including medical research, disease surveillance, clinical trials recruitment and service management, all of which are broadly considered to be secondary purposes to support the primary goal of providing care. Lower levels of de-identification often allow for more useful information to be released where data items are blurred or masked, and where identifiable data items are expressed as a pseudonym using a process referred to as pseudonymisation, which supports linkage between datasets as well as providing an indication of what the identifying data represent. A policy framework should state which must be applied and when, and these details depend on ethical reviews of the research projects and also need to incorporate details of safe working practice that is compliant with UK legislation and Department of Health guidelines.

1.3. Sharing Healthcare Records for Research

Healthcare providers across the world collect detailed information about people to manage their health needs and keep a record of their health and treatment. With a background of information governance requirements and increasing anxiety about information use and protection, the clinical profession has in its recent history recognised that healthcare delivery is a shared endeavour between different clinical teams and that digitising these records is an effective way of providing the information they each need to run healthcare services cost effectively and perform their duties correctly, effectively and in a timely fashion (Hillestad et al., 2005,

Kalra, 2002, PricewaterhouseCoopers LLP, 2013, Delaney, 2009). The profession and governments have also become more aware and supportive of patient empowerment through providing them access to their healthcare information and engaging them with their own health management (Sands and Wald, 2014, Her Majesty's Sationary Office, 2012, Hannan and Webber, 2007, Hannan, 2010).

A series of Electronic Healthcare Record (EHR) standards described in section 4.4.1 have been developed to ensure that a consistent view and understanding of the information held within records is communicated between clinicians and increasingly patients. An additional benefit of these standards and digitisation projects is that a wealth of detailed records are progressively being collected electronically and are more readily available to help provide care services and the activities that support it, including clinical research (Daniel and Choquet, 2014, Delaney et al., 2012). Chapter 4 discusses the healthcare information management strategy and funder backing for sharing healthcare records to provide effective healthcare services and support clinical research.

UK governments, which fund and run the NHS through the Department of Health (Department of Health, 2014b) have continuously pledged significant investment into supplying national scale Information Technology (IT) systems and infrastructure, identified as key to manage the information requirements and help improve efficiency, quality and effectiveness in NHS delivered care by supporting shared system and information use. Whilst the larger scale EHR system centralisation projects have been identified as over ambitious, unachievable and poor value for money (National Audit Office, 2011), care organisations within the UK are expected to maintain responsibility for their IT systems and continue sharing information to achieve care goals, improve services, ensure that these organisations can report on and get reimbursed for care provision and support research activities and collaborations with the pharmaceutical industry.

The focus of this thesis is on the use of healthcare records to support clinical research. The use of healthcare records in research is a compelling reason to share them: these range from population and cohort based studies that rely on the availability of hundreds of thousands of people's records to personalised medicine projects that focus on individuals. Both cases have enjoyed increased support in

legislation described in section 3.1.2 and collaborations across national and international boundaries continue to be funded. The last decade has seen significant investment by research councils into research computing facilities, all of which are focussed on the objective of being able to share resources and expertise to help answer research questions at a both population scale and to support personalised care for the individual patient, which are discussed in sections 4.1 and 4.2.

Research uses are a particularly interesting case: the NHS, Department of Health and clinical community do not regard research as a primary use of healthcare records (Department of Health, 2003) but recognise the value of research to care delivery (Department of Health, 2013b, Academy of Medical Sciences, 2006, Thomas and Walport, 2008) and expect records to be protected in line with primary use scenarios subject to their framework of information governance (see section 3.2 for a fuller discussion). There is some evidence of support for using healthcare records for research purposes from patients and clinical professionals as discussed in section 3.3, and whilst a recent report by the Royal Statistical Society demonstrates support for research uses from the UK public (Royal Statistical Society, 2014), it also highlights a "data trust deficit" where trust in institutions to use data appropriately is lower than trust in them in general. It is also clear that there are growing expectations with regard to protection and transparent communication about the purposes of that research, privacy anxieties (Stevenson et al., 2013, Barrett et al., 2006, Jamal et al., 2013) and some less supportive attitudes internationally (Whiddett et al., 2006).

1.4. Research Problem and Motivations for the Work

The Department of Health and NHS mandate a set of legislative and standards compliance requirements and have placed a responsibility on organisations to develop their own policies and remain compliant of those information governance stipulations. Focusing on the context of research institutions, there is a wealth of existing general institution scale policies, those that apply to individual projects or in some cases, little or any of these examples. Additionally, various information governance reviews (Department of Health, 2013b, Health and Social Care

Information Centre HSCIC, 2013) show an increasing expectation for organisations that process healthcare records to gain accreditation ether via the Information Governance Toolkit (IGT) (Department of Health, 2014a) or increasingly, independent ISO 27001 certification bodies. These must be combined with existing governance guidelines and data sharing agreements between collaborating organisations that required them to engage in a set of complex activities relating to a series of requirements that are not well understood by the majority of people engaged in research.

There has been no comprehensive set of requirements for determining the kind of information that is needed in the policies, a standard structure or guidance on how to prepare them incorporating all of these various stipulated items, or how to provide them to users and refine them to computer processable heuristics beyond involving human interpretation and interaction. Stipulations handed down by statutory bodies have required interpretation and deployment in different working environments, or have to be applied within a wider generic policy framework: policies may not be complete, working practice has to be guided by whatever is available, and it has been shown that users have a lack of understanding of their overarching legal and procedural basis (Becker, 2007, de Lusignan et al., 2007).

Through multiple clinical informatics research projects, engagement with stakeholders, interactions at conferences and the literature review discussed in Chapters 3, 4 and 5, the author has found that members of the research community are confronted with a number of policies that they must become familiar with, which are spread throughout websites, paper documents and often manifest in working practice that is not documented, leading to a tacit knowledge and expertise that is not reliably shared or practised, if at all. The problem is exacerbated by the nature of policies and data sharing agreements, which are invariably specified in a narrative structure to aid human understanding. This understanding relies on a consistent human interpretation of the narrative so that policy stipulations can be applied in practice, either by refinement to computable rules that protect information systems or to guide human behaviour, but understanding is not always achieved and human interpretations are not

consistent. Information governance is managed in an ad hoc, reactive and inconsistent fashion, leading to divergent working practice, where stipulations are either inconsistently interpreted or ignored.

Software configuration is also key to applying the appropriate controls for protecting information. Formal specifications exist to aid the process of specifying computable policy items, but these are not designed for human readability and are hard for people to author and interpret. Whilst computable systems are available for controlling access to resources they have to assign users to general groups based on their role and an arbitrary information sensitivity assignment, have been recognised as insufficient to meet required protection and do not integrate with wider security management, placing a greater reliance on people to behave correctly and appropriately. The literature recognises that many of the systems that have been built to help protect privacy have only been tested in laboratory environments and have not been deployed for industry use in actual projects. In practice projects and users must still interpret narrative policy and apply controls that they deem to be appropriate. The controls may involve configuration of software tools or involve users checking their own behaviour, for example if and when using USB drives to transfer information. A key solution for resolving issues regarding consistent understanding and interpretation is a comprehensive review of what information is needed to adequately inform policy definition and interpretation and how best to present and make it available to users to encourage a shared and consistent understanding of what is required of them.

1.5. Thesis of Research

The EHR is progressively enabling a consistent understanding and interpretation between a variety of different users through the adoption of standardised information models and constraints on the use of those models, which represent the structure of information within a domain of use and the relationships between data items that form part of that information. Concepts represented by this structured information are defined using a series of internationally agreed terms, where the relationships between the concepts that the constrained information models represent may be modelled by ontologies. Since information can be

represented and consistently understood for a variety of purposes, including to help inform decision making, automatically configure systems and provide necessary information to influence user behaviour, this raises the questions: can a knowledge management framework clarify the legal, ethical and good practice requirements for people when they specify policy and handle healthcare information for clinical research? Can it encourage a consistent understanding and interpretation of these information governance requirements whilst reducing the need to manually configure information system security components by automating the process of narrative policy refinement to computable settings?

The literature challenges the healthcare information management field to resolve the issue of comprehensively managing information governance for shared electronic healthcare information. Whilst the lack of a commonly agreed terminology within the domain of information governance makes the use of ontologies difficult to explore and establish, a knowledge management framework as used to implement systems that manage the sharing of electronic healthcare records still proposes a solution to consistently store information in a structure that can be shared and commonly understood. This kind of framework is constructed of a suite of software components designed to handle information from the point that it is entered by users and persisted in a database or other structured, static state through to the retrieval of that information by other systems and users, providing them with requisite knowledge to support their working practice and decision making needs. This approach relies on semantic consistency so that users can share a common understanding and interpretation of the information that is presented to them.

This work has established the knowledge requirements for information security, which have been established by reviewing relevant legislation, approvals bodies, international standards and good practice guidelines that exist as part of healthcare services to answer this question. This review has been supported by observation of working practice when constructing and using secondary use data repositories and engagement with stakeholders, particularly around issues of privacy, confidentiality and ethics in sharing healthcare information. These requirements have been analysed and have led to

the design and implementation of an information security knowledge model and a policy management application to specify and disseminate policy controls that aid human understanding and system configuration. These components form the proposed knowledge management framework, which has been evaluated as described in Chapter 8 to validate its use as a proof of concept solution and to test the three hypotheses that have been proposed earlier in this chapter.

1.6. Thesis Structure

This thesis is composed of twelve chapters structured as follows:

- Chapter 2 describes the materials and methods used to determine the information requirements, use these to develop the knowledge management framework and to assess its use by live participants;
- Chapter 3 provides background material that describes information governance frameworks in the UK, including the rights to privacy, duty of confidentiality, need for data protection and application of suitable controls using information security. It also describes the issues that these raise for using healthcare records in medical research;
- Chapter 4 describes developments and facilities for sharing healthcare records to support clinical care and research and the international electronic healthcare records standards that have been developed to support the healthcare improvement strategies;
- Chapter 5 describes the case study exemplars, learning outcomes and contributions to the research work;
- Chapter 6 presents the requirements and design of the proposed knowledge management framework to manage information governance;
- Chapter 7 describes the implementation of the knowledge management framework that has been developed, focusing on the EHR standards and research upon which it is based, the technologies used to implement it and the tool that they have generated;

- Chapter 8 describes the approach taken to evaluate the framework and how the evaluations were conducted, providing the analysis obtained from the evaluations to test the hypotheses and evaluate the approach as a proof of concept implementation;
- Chapter 9 critically appraises the work and discusses its limitations.
- Chapter 10 describes proposed further work and provides the conclusions to the research work;
- Chapter 11 provides a bibliography of citations and references.

A set of appendices has been provided at the end of the thesis, which is referred to throughout this work. The next chapter describes the materials and methods that have been used to conduct the research work, develop, implement and evaluate the proposed knowledge management framework.

Chapter 2. Materials and Methods

This thesis has explored the area of information governance for electronic healthcare records when used for clinical research. The aim of the work has been to design and develop a knowledge management framework, which has been proposed as a way to simplify the process of developing information security policies and to support people in understanding their responsibilities when managing the information derived from these EHRs. The focus of this work has therefore been around software systems engineering and understanding the complexities of managing healthcare information for use in the clinical research context in accordance with legal, ethical and security requirements.

The materials and methods used to achieve these goals focused on identifying, reviewing and understanding requirements sources for the proposed framework. They also focused on developing an understanding of current practice in managing healthcare research information assets and how to handle the issues that have been identified in this work. This would provide the basis for pursuing the software systems engineering goals, which needed a well recognised, structured and tested approach to develop the knowledge model and resulting frameworks for evaluation. The Universal Software Development Process (USDP) (Jacobson et al., 1999) provided this approach, which proposes the specification of requirements to help the design and development of a software system so that it could be implemented and tested.

A significant source of the requirements for information governance have come from data protection and confidentiality legislation, in addition to human rights and health service governing law that provides a legal basis for managing information sharing. This legislation has informed a series of guideline documents that have been developed by the UK Department of Health to guide healthcare organisations on their responsibilities when processing healthcare records. These guidelines have also been informed by the development of national and international standards for information security management. The requirements sources therefore include legislation, guidelines and standards. These sources had to be reviewed and considered in the context of peer reviewed literature on their

use and effectiveness as well as their use in practice when running research projects that relied on information derived from healthcare records.

A combination of literature, legislative and good practice guideline reviews and case studies of existing research projects were therefore identified as a means to explore and understand common practice, support mechanisms and peer reviewed assessment of the problems with the existing approaches. This helped to illustrate the requirements sources and specify them, analyse them and design the knowledge model and management framework. It also helped to determine the optimal deployment method and establish a meaningful evaluation of the hypotheses.

The majority of the research work has been carried out at the UCL Centre for Health Informatics and Multiprofessional Education (CHIME). It has been accomplished through ongoing research and investigation carried out during several UK and European Commission funded projects where the development of clinical data warehouses, reuse of EHRs and understanding the information governance challenges was key to developing new infrastructures and answering clinical and health informatics research questions in a safe, legally compliant environment.

2.1. Study Design

Figure 1 illustrates the process by which the work of this thesis has evolved through a succession of research and clinical system deployment projects, each involving the use of clinical information for care, research or other secondary purposes and therefore facing the security and confidentiality protection challenges addressed in this thesis. Through this process the work has been informed by an improved understanding of the challenges it needs to meet and ways to address these have evolved, as reflected in the progression of foci. The successive foci of the research, shown in white, have guided the generation of goals that are needed to explore the research area, shown in yellow. These goals have been achieved using the listed methods and given rise to a series of outcomes upon which the thesis can be evaluated.

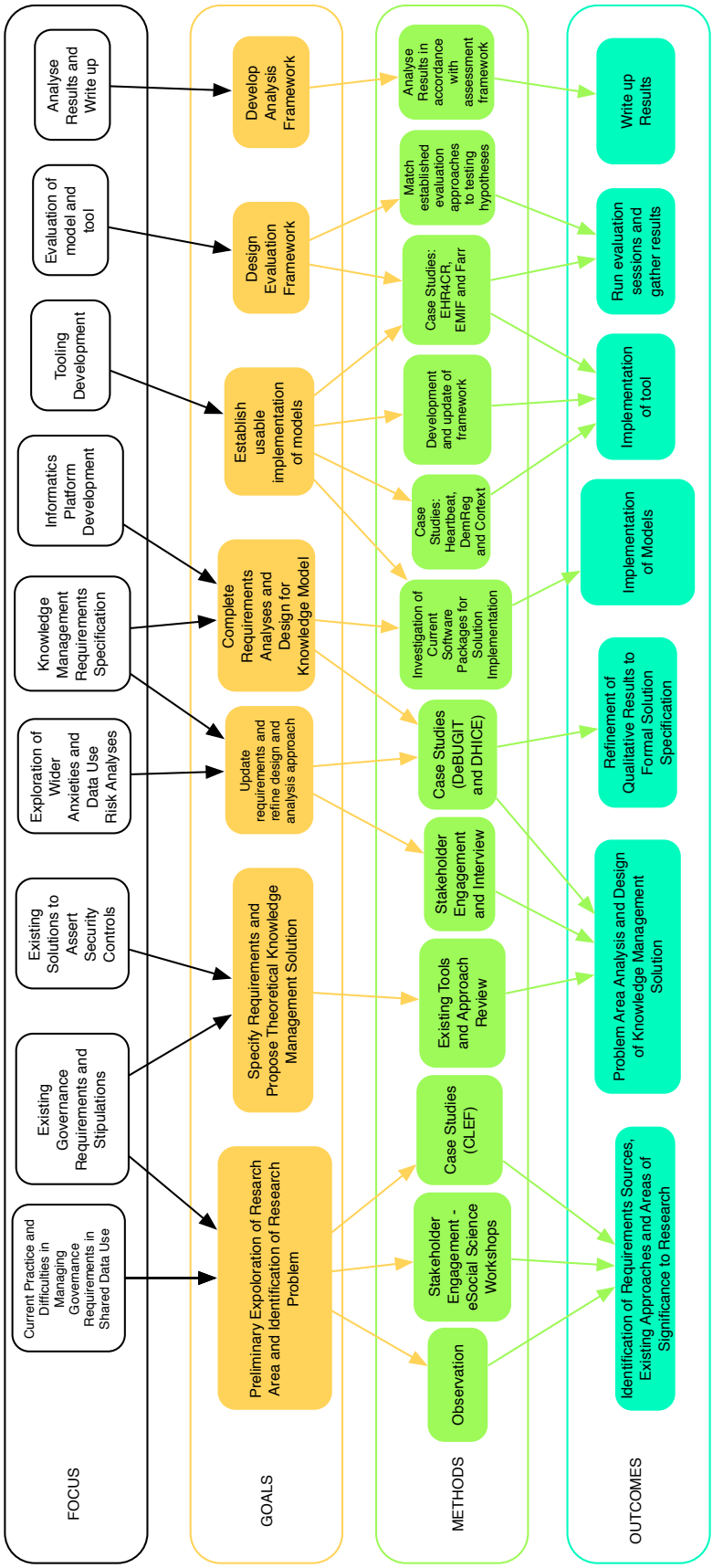


Figure 1: Study Design

The approach has been iterative, which allowed for a steady progression in gathering requirements, designing and developing the knowledge management framework to evaluating the solution, reiterating over previous steps when attempts at requirements analysis, design or implementation prompted a review.

By using the USDP to lead the software development, it was possible to perform a review of the relevant literature within the research domain and incorporate learning outcomes into the development process. Working on EHR information sharing projects and observing their use in practice allowed the author to become more familiar with the legislative and good practice guidelines and their effects on and interpretation in working practice. This provided a structured approach to developing the software as well as pursing the research work and exploring the research area through the literature reviews and case studies. The following sections discuss the literature reviews, case studies with an overview of their contribution to the research work. It also describes the other methods and outcomes as shown in Figure 1, including the development and implementation of the knowledge management framework and its evaluation using live participants to test the hypotheses and its effectiveness as a proof of concept solution to managing information governance requirements.

2.2. Literature Review

The literature review has been conducted by making use of a variety of bibliographical resources. The availability of literature online and through UCL and University of London Library Services has meant that it has been possible to examine the literature using indexing services, including:

- Citeseer
- PubMed
- Scopius
- JStore
- Discovery @UCL
- Google Scholar

This research is multidisciplinary in its nature, so publications in the fields of medical, biomedical, legal, computer science and engineering, human computer interaction, ethics and research methods were reviewed. These included journal articles and conference proceedings found in:

- The British Medical Journal (BMJ)
- The Journal of the American Medical Informatics Association (JAMIA)
- The International Journal of Medical Informatics (IJMI)
- Ethics Review
- BioMed Research International
- BioMed Central Bioinformatics
- BioMed Central Infectious Diseases
- BioMed Central Medical Research Methodology
- BioMed Central Medical Informatics and Decision Making
- European Journal of Epidemiology
- Future Generation Computer Systems
- IEEE Transactions on Knowledge and Data Engineering
- Informatics in Primary Care
- International Journal of Healthcare Information Systems and Informatics (IJHISI)
- Journal of Clinical Oncology
- Journal of Biomedical Informatics
- Methods of Information in Medicine
- Nature
- Social Science & Medicine
- Studies in Health Technology and Informatics
- Telemedicine Journal and e-Health
- Transactions on Data Privacy

A series of search terms was developed based upon commonly used terms found in standards, guidelines, press articles, commentary, conference presentations and

keywords found within reviewed papers (where available). The search terms included:

- Privacy Research
- Privacy Research Health
- Privacy Research Biomedicine
- Confidentiality / Duty of Confidence
- Confidentiality Research
- Confidentiality Research Health
- Confidentiality Research Biomedicine
- Information Security
- Information Security Research
- Information Security Research Health
- Information Security Research Biomedicine
- Information Governance
- Information Governance Research
- Information Governance Research Health
- Information Governance Research Biomedicine
- Knowledge Management
- Knowledge Management Privacy
- Knowledge Management Confidentiality
- Knowledge Management Duty of Confidence
- Knowledge Management Information Security
- Knowledge Management Information Governance
- Privacy Concerns Health Information Sharing
- Confiden* Concerns Health Information Sharing
- Information Security Concerns Health Information Sharing
- Information Governance Concerns Health Information Sharing
- Legal Concerns Health Information Sharing
- Concerns Health Information Sharing Health and Social Care Act
- Concerns Health and Social Care Information Centre

- Concerns Section 251 Exemption
- Research Ethics
- Research Ethics Committees
- Consent clinical medical research
- Consent clinical / medical research opt-in opt-out

Manual searches were also conducted at UCL and the British Library where materials were not available online. In all cases, the author reviewed the titles of search results and their abstracts to assess relevance using a set of inclusion and exclusion criteria. The inclusion criteria were publications that focused on managing information security or governance requirements for shared clinical care or research purposes, methods and software tools for protecting privacy and upholding confidentiality (for example de-identification techniques and / or access control), attitudes of professionals and patients to information governance and its effectiveness and concerns about the sharing of healthcare records for care or other purposes.

Publications relating to the modelling or formal representations of security policies were also included in the review, as well as those that related to the usability of tooling to aid policy specification and interpretation. International articles were included, and no cut off date was used. Exclusion criteria were publications that had no applicability to healthcare record management, areas of healthcare provision that were not relevant (including surgical and nursing techniques and technologies that were designed for patient self management of chronic conditions where there was no relevance to information governance). Articles that were not relevant were rejected, and those that were relevant were read in detail. Of approximately 15,000 search results, about 250 articles were relevant to this work. This helped to identify themes and categories for the literature, where an overview of these themes is discussed below. The outcomes of the literature review are presented and discussed in detail in Chapters 3 and 4.

Several themes were apparent across the literature, which fell into categories around societal expectations and concerns founded on legal, procedural and ethical requirements, which are discussed in detail in Chapter 3. The categories

included a consideration of issues with the technical implementations that existing and pioneering information governance methods have introduced and their effectiveness in the area of more shareable EHR systems and uses of the information they contain. Other categories include the usefulness and importance of detailed records for research and the engineering of shareable and standardised EHRs with associated tooling for improving care outcomes and the reusability of EHRs for other purposes. These categories are described in detail in Chapter 4.

The most frequently occurring theme related to privacy and confidentiality of EHRs. Approximately one hundred publications considered the confidentiality of EHRs and about ninety considered privacy, with some overlap between the two. These papers considered the risks and challenges posed by using a more shareable, accessible electronic medium to handle the records and specific challenges posed by research, including epidemiological or genomics research. Several articles related to public and professional attitudes to sharing healthcare records for research purposes, indicating some support overall for medical research using EHRs collected during routine care, though with expectations about being contacted to be asked for consent and informed about the research that was underway. This was balanced by publications regarding public concerns and issues over larger scale projects, including the Summary Care Record and Care.data, where anxieties were raised over the how the confidentiality of the information would be protected, who would have access to it and whether it would be "sold" to industry.

The legal ramifications of privacy and confidentiality for some of the larger scale healthcare record management projects have featured in the literature. There have been several debates surrounding the need to get consent from individuals, particularly for nationwide, population scale initiatives. Several papers have been published regarding the possible approach of asking people for their consent to participate in these initiatives or whether people should be asked to opt-out of them if they do not want to participate. The discussions also focussed on whether the opt-in or opt-out approach could be regarded as explicit, meaningful consent, and how it affects participation numbers.

Approximately twenty of the articles focused on de-identification strategies for different areas of clinical practice and research. These either worked to anonymise the records or to provide pseudonyms and enable linkage between datasets whilst claiming to protect confidentiality. These publications included discussions on the reliability of the de-identification strategies and the remaining risks of re-identification associated with anonymised or pseudonymised datasets.

About thirty articles discussed record linkage itself in the context of improving the completeness of records for care purposes and for providing richer data sources for research purposes. There have been more publications in the last three years around the possibilities of linking EHRs to the genome sequences of individuals to consider possible patterns associated with treatment and genetic structures. These discussions have been considered with the risks of re-identification of de-identified records because a more detailed profile of an individual is established when records are linked, making them potentially easier to identify. The linkage publications consider privacy protecting strategies and the risks to confidentiality, proposing new strategies to protect identities and other information and scrutinising existing methods. Whilst there are clearly potential benefits of linking records and other data sources and there are some examples of the benefits of previous linkage works that have been identified, the case for linking healthcare records to social care and genetic sequences has not yet made as strong a case for the additional risks to participant identification, anonymity and confidentiality.

About fifty publications discussed research ethics, the role of ethics reviews and the committees that have been established in the UK to review the ethical basis for research projects. Of these publications, the themes that emerged included discussions around the trade off between individual rights and the good of society and the public, or even the wellbeing of that individual. Several publications reviewed the effectiveness and limitations of informed consent. The role and effectiveness of ethics committees themselves were also reviewed, discussing their applicability, their membership profile and the quality of their deliberations and rulings.

Most of the articles critiqued specific tools or approaches for technical solutions to maintaining confidentiality, privacy and security. Only a few of the publications alluded to the issues with current practice more specifically, linking them back to legal and procedural requirements and identifying that there was a lack of clarity and understanding of the legal basis for handling healthcare data and that these were not being clearly communicated. This was in addition to the issues with interpreting information security policy for enactment in the area of managing healthcare records, or identifying that protecting privacy and managing confidentiality was important and needed to be handled appropriately.

In each case, they have pointed to the need for a solution to this lack of clarity and issues surrounding understanding and interpretation, but have so far not provided any proposal for handling the situation. Additionally the literature has not provided a clear overall definition or scope of information governance for sensitive healthcare records in the context of care or research uses. It has also not proposed an integrated, holistic approach for managing information governance requirements, or offered an authoritative set of these requirements spanning the legal and good practice guidelines when enacted using established information security techniques and tailored for reasonable information processing.

The area of EHR design and development featured in about eighty publications, where the themes included several examples of semantic interoperability between systems and different standards. Others related to implementation according to standards and their compliance and querying the EHRs and development using particular technologies were also found. There were very few that described actual effects, benefits or otherwise on clinical care itself, a point noted in the literature.

In addition to EHRs, the technical security protection mechanisms for them also featured in the literature review. Approximately twenty of these related to access controls, with role and purpose based methods featuring in the examples. There were also several publications around privilege management within systems. Encryption, authentication and authorisation also featured as part of the larger scale infrastructures and privacy preserving linkage work. A description of the existing scripting methods that are used to configure access control,

authentication and authorisation systems was also discovered in the literature, as well as examples of formal specification of policy items.

The literature review provided a series of themes around the area of research uses of EHRs and some gaps in the literature have been identified and are discussed in more detail in Chapter 4. The literature review also helped to identify some of the key legislation surrounding the protection of EHRs and the people that they are held about. This also indicated the guidelines and international standards that were relevant to the area. The case studies described in the next section allowed the author to discover more examples of pertinent legislation, guidelines and standards, the organisations responsible for specifying the guidelines and delivering healthcare, their effects on working practice and other practical issues that they raised.

2.3. Research Project Case Studies

The case studies allowed the author to further explore the issues surrounding information governance whilst using healthcare records for clinical research. This included observation and understanding around the key legislation identified in the literature review in practice, how guidelines derived from that legislation informed working practice and how information security standards were deployed to aid working practice. This also helped to identify the organisations that were responsible for developing good working practice guidelines, legal compliance and the delivery of healthcare and funding clinical research. These observations helped to hone the understanding of the issues that were being raised and the difficulties that were encountered, allowing for a clearer articulation and appreciation of the problem of managing information governance in this context. Changes to legislation and good practice guidelines could also be observed both in the context of reviewing the literature and working on the case studies. The case studies also allowed the author to engage with researchers working within these projects, patient groups and other stakeholders to further understand how traditional approaches to information governance affected the people involved with research. The case studies are described in detail in Chapter 5.

The research work commenced in late 2004 with an assessment of the challenges facing the research community when seeking to gain access to and use sensitive healthcare records. This assessment was performed whilst working on the UK Medical Research Council (MRC) Clinical eScience Framework (CLEF) project, which aimed to develop a standards-compliant, secure clinical data warehouse and policy framework and highlighted issues of complexity, understanding and enactment of information governance management. During this work, the author developed, assessed and investigated the core requirements for establishing an information security policy and information needs. This helped to identify the research areas for the thesis and requirements sources from legislation, good practice guidelines and standards for the development of a comprehensive information governance framework that would guide research uses of electronic healthcare records.

The European Commission funded DebugIT project developed a clinical data repository that held infectious disease information derived from electronic healthcare records across several European countries. The author investigated the knowledge management requirements for establishing a policy framework that permitted data release to research partners across Europe. This enabled the review of further requirements sources that applied to the UK and European contexts and informed the evolution of the information governance framework and a generic policy template for research uses of information.

The Databases for HIV: Integration, Collaboration and Engagement (DHICE) Initiative offered an opportunity to observe working practice to protect information assets across a series of HIV cohort and surveillance studies operated by UCL, the MRC and Public Health England (then the Health Protection Agency) that obtained their data from multiple specialist treatment centres in the UK. This allowed us to further enrich the policy framework based on additional information governance requirements for disease surveillance studies, support secure linkage between the studies' datasets and provide a generic framework to the projects so that they could develop their internal policies. It also offered the opportunity to design a knowledge model called the *Secutype* that would provide both a human readable and a computable representation of policy clauses according to the

governance framework requirements. The work led to the development of generic constraint model known as the *Pattern*, which evolved from the EHR Standard EN ISO 13606 concept of the Archetype constraint model (Beale, 2002, British Standards Institute, 2007b) and used the same standard's specification of a Reference model representing a record structure to allow the expression of both clinical and information governance knowledge models, through which a common representation of information security policy components could be provided. The development of the Pattern and use of the standard models are described in more detail in Chapter 6.

In parallel to the research projects listed above, an electronic health records development team based at CHIME, of which the author is a part, has developed and deployed systems for cardiovascular shared care (Heartbeat), dementia clinical trials recruitment (DemReg) and dementia clinical research (Cortext). The Heartbeat, DemReg and Cortext projects involved the development of live clinical systems and dementia registries. These projects used the Pattern model to generate a database schema and clinical screens within a web application architecture to achieve a clinical information system that was compliant with the EN ISO 13606 standard. This provided an opportunity to develop the framework to achieve these goals for implementing the Secutype model as Patterns, which guided the generation of the information governance advisory tool as a web application.

The EHR4CR, EMIF and development of the Farr Institute have involved the development of larger scale, nationwide or European data repositories that make use of newer data storage facilities for larger patient populations, with aims to link with genomics and social care information. The infrastructures sit within "data safe havens" that are established based upon current information governance guidelines and certification requirements. This provided an opportunity to explore a new paradigm of storing identifiable data in research repositories with a more critical requirement for effective protection measures.

Each of the case studies helped to illustrate the issues with managing information governance and to provide more examples of requirements sources to develop the solution. By using the USDP, they also allowed the author to refine and

specify these requirements in accordance with observed working practice, design the knowledge management solution and develop it. This provided the basis for implementing a solution and piloting it, as described in the next section.

2.4. Implementation of Information Models and Tooling

Once the requirements specification and analysis had reached a level of completion, the design and development of the knowledge management framework could commence as the work progressed through the case study evaluation and the literature review continued. The use of the USDP provided a series of targets that needed to be reached for the development process, which could be managed iteratively as the case studies proceeded, changes to legislation came into effect and any other factors could be incorporated into the development process. Once the specifications had reached a level of maturity, the first versions of the systems could be implemented. This implementation was based upon the Secutype knowledge model and its implementation using the Pattern, which is described in Chapter 7.

The implementation of the Pattern and Secutype models has been possible through collaboration with a team of developers at CHIME working on a knowledge management framework called the Clinic Manager 3 (C3). This framework has been developed to implement clinical record systems as web applications used in practice and for disease research repositories as used in the Heartbeat and Cortext projects. The C3 framework used to generate the tooling has been under development during the progression of this work. The author has helped to implement the framework so that it can generate a web application for policy editing and review based upon instances of the Secutypes. This has resulted in the generation of user screens and a database schema that could be presented to participants for use in the evaluations. By using this approach, the tool has a suite of auditing and version management capabilities in line with the requirements identified in Chapter 6. The implementation work created the proposed knowledge management solution, allowing an opportunity to test, evolve and validate the Secutype model design through a series of incremental steps and pilot evaluations. This would bring the development work to a point where it could be

used to test the hypotheses and validate the knowledge management solution as a proof of concept solution to managing information governance requirements when handling electronic healthcare records for research.

2.5. Evaluation Approach and Hypothesis Testing

The focus of this work has been to determine the effects, benefits and detriments of using the Secutype based web application tool on user decisions and behaviour when processing healthcare records used for research purposes. The evaluation of a web tool traditionally involves a series of assessments that focus on user experience and utility for a given task, within the context of the user's specific working environment and how it affects wider team members and the organisation within which they work. In the case of EHRs used for research, research institutions have to mitigate risks to their business when processing these records, and (ideally) much of this is managed through policy-based controls. Any evaluation of a knowledge-managed approach must therefore include a representation of organisational stipulations, informed by a governance framework that is consistent with legal and ethical expectations within which it sits. Having a clear specification of the expected outcomes is also important, as well as a means to measure participant behaviour when using the tool and their own opinions of the experience and how they view its utility in common practice.

Measuring the effects of a tool built to support user decisions about sharing healthcare records used in research relied upon an understanding of the information governance expectations and stipulations. These focus on a legal compliance established through expression in policy as a general framework with specific details that have been established through risk assessments to guide day-to-day working practice. Assessing the effects of a new means for users to author instructions and deliver these instructions for effective guidance must measure these effects on participants. Whilst no comprehensive evaluation framework exists to assess tool compliance with information governance requirements, the focus of the evaluations had to test the hypotheses outlined in section 1.3 and provide:

1. a clear, proven assessment of the completeness of the knowledge model and the information that it provided participants;
2. a measure of participant understanding and successful outcomes when they used the tool for authoring policy, review and decision making;
3. a means to measure any unexpected or emergent participant behaviour when reconciling their practice with conformance to authored policies;
4. what the participants' views were in using the tool, whether they felt it would clarify what was expected of them in terms of information governance compliance and whether they would find it useful in their work.

The evaluation approach involved the execution of two stages. The first stage sought to contact potential participants and request that they provide excerpts from information security policies, data sharing agreements or details of working practice from their organisations when handling sensitive information. Once participants had agreed to participate and provided the excerpts, the second stage could commence: this involved a series of three hour sessions where participants were invited to use *keibi* in a set of experiments that would test their ability to author policies and guide their behaviour in a series of exercises carried out in a laboratory setting. The stages are described in this section. This evaluation approach was piloted in March 2013, where a series of minor improvements were made to *keibi* and the approach itself, and the main evaluations were completed throughout December 2013 and January 2014. Appendix 7 provides details of how the pilot evaluations informed the improvements.

The first stage involved the identification of participants and collection of candidate policies. The goal was to collect materials to run the experiments in the second stage. It achieved this by preparing a list of appropriate participants, approaching them with an invitation to participate. If they accepted, they were asked to provide excerpts of policy, data sharing agreements or working practice details that they use when processing sensitive information. Figure 2 provides an overview of each step in the process of executing stage one:

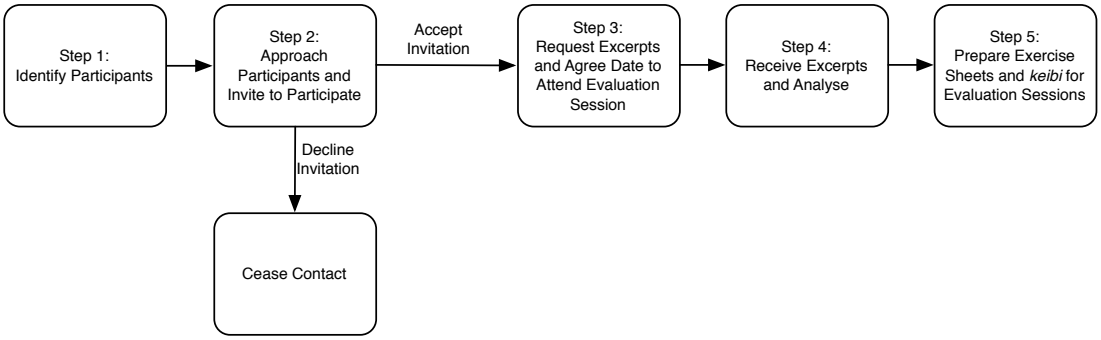


Figure 2: Steps Executed in Evaluation Stage One

The first stage involved the selection of appropriate potential participants that could be approached to participate. The author opted to scope the evaluation as an ethnographic study in order to gather a set of results that represented the experiences of the range of different skills and experiences within the context of research uses of medical information. Participants were therefore selected from research institutions including University College London (UCL), specifically the School of Life and Medical Sciences (SLMS) and Faculty of Engineering, where research computing facilities and the data safe haven are being established. Participants held affiliations with a number of research institutions and NHS trusts, as described in section 8.1. Selection criteria for participants included: experience of working in the area of medical research with a focus on healthcare information processing; association with a research institution; and either involvement in information security and governance policy development and / or execution, or directly handling sensitive information for research purposes, or both. Potential participants would include professionals from several fields, including clinicians, computer scientists, IT Professionals, information governance and / or security experts, and statisticians or epidemiologists, where they would sometimes occupy more than one of these professions. A proposal was put to the UCL Ethics Committee to determine whether the evaluations would be exempt from ethical review on the basis that participation was unlikely to cause participants harm, would not be used to judge their working performance and they would remain anonymous. The proposal and the response from UCL Ethics confirming exemption are available in Appendix 8.

The next step involved inviting participants through email invitation, which included a brief introductory document to *keibi* and what participation entailed (see Appendix 9 for the invitation emails and the attached introductory documents). On accepting the invitation, the participants were assigned a participant number and were asked by email to provide up to three clauses from an information security policy of their choosing that involve the handling and sharing of information from their organisation, up to three clauses from any data sharing agreements that they have entered into prior to the session and to document up to three examples of common working practice when managing security, which are not to their knowledge specified in writing, policy or sharing agreement. Guidelines for submitting these excerpts and a spreadsheet for returning them were provided in the email; these can be found in Chapter 10.

Participants were asked to provide their own excerpts because it was not reasonable to make any assumptions about how *keibi* might be used in practice: one possibility was that it would be used to document existing policy or working practice. By asking participants to provide their own examples of existing stipulations, this not only provided an opportunity to see how well *keibi* guided them in updating existing documentation, but it also helped to ensure an impartial evaluation of *keibi* in stage two of the evaluations to see how well the tool performed with policies developed by people other than the author, and used in actual practice.

Policy and Data Sharing Agreement excerpts and examples of working practice were sought because it was not clear whether the different divisions and projects within the SLMS had any policies, data sharing agreements or indeed established working practice in place, though it was a reasonably safe assumption that at least one of these would be available given the facilities and data processing that occurs. Participants were sent a spread-sheet where they could enter the required details, which included the excerpts themselves, their type (policy, data sharing agreement or working practice), their source, the intended outcome for reader behaviour, whether they were intended for human readership, software configuration or both. Participants were also provided a brief guideline document for completing the spreadsheet (see Appendix 10). Participants were given no less than three

weeks to provide the excerpts, and were asked to provide the excerpts no later than four days prior to the evaluation date, which was agreed when their participation was confirmed and they were asked to provide the spread-sheets.

The fourth step involved a review and analysis of the provided excerpts. This was to ensure consistency across the different evaluation sessions and to make sure that that there was a basis for comparison across the different sessions, as well as the presence of at least one excerpt that was intended for software configuration. The excerpts had to be rich in detail to make sure that the features of *keibi* were fully tested. The selection was also based on how much standardisation would need to occur to provide consistency and remove ambiguity from the excerpts; excerpts that required less or no standardisation were considered more suitable because they represented the excerpt in its most original form without interference by the review process. Once appropriate excerpts had been selected, Step Five commenced, where exercise sheets for the evaluations could be prepared as described in section 8.2 and *keibi* could be configured with the appropriate user accounts for the evaluations sessions. The exercise sheets can be found in Appendix 13 and are described in more detail with expected answers in Appendix 14. *keibi's* configuration entailed providing user accounts for each participant, as well as Use Contexts (described in section 6.7.2) where participants could enter the details required by the question sheets. Participant numbers were used to form the basis of the account details - no identifiable information was used in their accounts.

The second stage comprised three experiments to evaluate the knowledge management framework, which are described in section 8.3. The first experiment involved the use of *keibi* in practice for the participants to author a selection of the supplied policy excerpts. The second experiment involved the use of *keibi* in practice for reviewing and using policy details for participants to determine how to handle sensitive information assets in given scenarios. The third experiment involved the completion of a user satisfaction questionnaire and focus group discussions about using *keibi* to further evaluate the knowledge management based approach as a proof of concept solution. The first, second and third experiments were run over a three-hour workshop in a laboratory environment

over five separate sessions, where participants attended and ran through the experiments using participant specific question sheets. The question sheets provided the policy excerpts that participants used to author policy items in the first experiment, provided questions and a space to answer them for the second experiment, and included the user satisfaction questionnaire, which formed part of the third experiment.

2.6. Analysis of Evaluation Results

Once the evaluations were completed, the responses entered into *keibi* for the first experiment were reviewed and those from the exercise sheets in the second experiment and the user satisfaction questionnaire and group discussion in the third experiment were transcribed. Section 8.5 describes the method of analysis in detail. In the case of the *keibi* policy items that had been authored by the participants, each one was assessed to see whether participants had understood what they were supposed to enter by comparing them with the submitted expected outcomes. The entered items were also assessed to see which of the Secutype knowledge model classes were used to express the required policy items. The entered items were also scored according to the EN ISO 13606 Reference Model classes that were used to structure the Secutype classes that were instantiated.

The transcribed answers gathered in Question 2 were totalled to determine the number of answers that were correct based upon the expected responses derived from participant excerpt contributions, as well as those that were incorrect and omitted. In both experiment one and two, assessments were made where participants exceeded expectations or showed that they made unexpected responses based on their own expertise. The results from the user satisfaction questionnaire, which included space for participants to make their own comments, and from the group discussion provided additional material in the form of their own feedback. This helped to provide further details to explain some of the responses in the first two experiments, as well as establish evidence to support the use of the knowledge management approach as a proof of concept solution by gathering participant opinions and views based on their experience of using *keibi*.

The results from the first and second experiments were aggregated to draw conclusions about whether the hypotheses were supported based on participant responses. The results of the third experiment were used to develop an a more complete evidence basis for validating the approach as a proof of concept by combining the feedback from participants with the results of analysing their responses. From these results, it was possible to draw conclusions about the usefulness of the approach, whether it was validated as a proof of concept and to discover where further work would be appropriate to understand the optimal approach for managing information governance in clinical research.

Chapter 3. What is Expected of Best Working Practice in Research?

The handling of identifiable healthcare information is subject to a series of restrictions and societal expectations, which are protected by law, supported by guidelines for working practice and may be enacted by techniques that are defined by international standards. There is an international trend to meet these expectations and enforce these restrictions and, whilst there is no one resource in any country that authoritatively and comprehensively defines legal requirements and adherence to good practice when processing healthcare records for any purpose, be it clinical care or research, the restrictions and expectations uphold the individual's right to keep details about herself or himself private, the medical profession's duty of confidentiality and data protection. In Europe these are largely defined by national Data Protection legislation, which is upheld and achieved through guidelines, systems and tooling that is defined and established according to information security standards to inform good working practice. The sharing of healthcare information for research, be it large scale epidemiological research, preparation of aggregated data sets or the use of identifiable information in smaller scale cohort studies, each represent challenges for the protection of privacy, duty of confidentiality and the expectations of patients and research participants.

In the UK there is general agreement that these requirements, expectations and challenges are to be met through effective *information governance*. Several groups responsible for running healthcare services, commissioning and managing health and / or social care information have been established. For example, the Health and Social Care Information Centre (HSCIC), the "national provider of information, data and IT systems for health and social care" (Health and Social Care Information Centre, 2014a) defines information governance as "…(ensuring) necessary safeguards for, and appropriate use of, patient and personal information. Key areas are information policy for health and social care, IG standards for systems and development of guidance for NHS and partner organisations." (Health and Social

Care Information Centre (HSCIC), 2014a). NHS England, the body responsible for improving healthcare outcomes for people accessing NHS services in England and commissioning of these services, refers to Acts of Law that establish the need to protect data and allows sharing of healthcare information for essential secondary uses to support the provision of healthcare: "the legal framework governing the use of personal confidential data in health care is complex. It includes the NHS Act 2006, the Health and Social Care Act 2012, the Data Protection Act, and the Human Rights Act." The Health Research Authority (HRA), established in December 2011 to "…promote and protect the interests of patients in health research and to streamline the regulation of research…" (Health Research Authority, 2014a) defines information governance by grouping a set of resources pertaining to personal data use in research, research databases based on data protection legislation (Health Research Authority, 2014b). One resource it alludes to is the Information Governance Toolkit (IGT), described in more detail in section 3.2.2: the IGT describes information governance as being "…to do with the way organisations 'process' or handle information. It covers personal information, i.e. that relating to patients/service users and employees, and corporate information, e.g. financial and accounting records…" (Department of Health, 2014a).

The literature has started to evolve the understanding and scope of information governance to include legislative and ethical requirements as well as information security techniques. It proposes that the required interpretation and enactment should be within a framework of principles rather than rules, which can more effectively inform sharing decisions based on appropriately guided risk assessments. Laurie et al. proposed that good governance must take into account the requirements of the information processing and be handled on the basis of well established principles which allow flexibility in interpreting legal requirements as opposed to rules, which have been identified as unnecessarily exclusive to managing healthcare information and inflexible to the point that decisions about reasonable sharing of information are not well informed. By using the principles approach, effective risk assessment can be informed, which is key to ensuring that people are engaged and empowered to make effective decisions on grating access and base this on an effective risk assessment in line with data protection, ethical

principles, and risk assessment and technique use around access control and de-identification, which represent elements of information security management (Laurie and Sethi, 2013).

There is a common agreement that information governance is based upon legal requirements and ethical guidelines, which are intended to guide safe working practice when processing healthcare information in the context of care or secondary uses. This working practice must be guided by effective risk assessment and is enacted by information security techniques as defined in international standards. The relationship between these elements of good governance and articulation of societal expectations have not been fully defined or explored. The expectations of good practice described in this chapter and case studies described in Chapter 5 illustrate the importance of stakeholder engagement for healthcare information reuse and protecting the relationship between the patient and their healthcare provider, which is the point that this information is recorded. The author therefore proposes in this thesis to consider information governance as the management of legal, ethical and good practice requirements using information security techniques to protect the relationship between the patient and health service provider by guiding people in safe, secure working practice when handling healthcare information.

This definition has been illustrated by adapting the poem by Rudyard Kipling *the Six Honest Serving Men*, namely six key questions Who, What, Where, When Why and How to provide some explanation of the requirements and expectations, and how they must work together to provide effective protection and guidance when using sensitive healthcare information for any purpose:

Why we must protect healthcare records:

- Confidentiality – the duty of the medical profession, foundation of its relationship with the patient and protected by the Common Law Duty of Confidence;
- Privacy – a right of every individual; this right must be upheld and is protected in statute (for example the Human Rights Act).

What we are doing to protect them:

- Data (and therefore Service and Stakeholder) Protection – enshrined in the Data Protection Act 1998 and its European equivalent the European Parliament and Council Directive 95/46/EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. The Data Protection Act provides a series of principles that define the individuals' rights when information about them is processed. These rights include non-maleficence, which is espoused by ethical principles for the practice of medicine and use of information for ethically approved research.

Who protects them, ***when***, ***where*** and ***how*** they are protected:

- Established by Information Security – the practicalities, authoritatively and comprehensively defined by the ISO 27000 series of security standards. This will involve the use of risk assessment to specify policy and information security management systems to guide people who handle and use healthcare records of how to behave with it, as well as purpose hardware and software that are designed to control access to and use of that information.

The author has used this approach as a teaching tool within UCL Medical School for undergraduate and postgraduate courses, which students have found helpful to illustrate how these elements work together in practice for managing access to, use and protection of electronic healthcare records.

This chapter explores these core elements, explains the frameworks that have been established to enact them within the UK and research context and the organisations that are responsible for ensuring their enactment and use in practice. This is combined with the results of the literature review related to these areas. It commences with a description of the legislation that is pertinent for data protection, privacy and confidentiality. This leads into a discussion around de-identification of patient records, the use of identifiable data and the requirements for and exemptions to managing consent. This is followed by a consideration of the ethical protections and risk management that is in place as a result of the legal mechanisms. The practicalities of information security management are then

discussed, guided by these requirements sources, which include a discussion around good practice guidelines that have been developed based on the standards and legislation. The issues with these approaches are then described, preceding an overview of societal attitudes and anxieties that have come developed around the sharing of sensitive records. The chapter concludes with a summary of the findings from the exploration and the literature review.

3.1. Legal Framework: Data Protection, Privacy and Confidentiality

Figure 3 depicts the pertinent legislation for the processing of healthcare records for care and research purposes. Legislation in the UK broadly mandates that information can only be released and used for specific purposes according to very strict stipulations that are designed to mitigate any risks of causing harm to the individual from whom it is collected. A core piece of legislation is the Data Protection Act 1998 (DPA)(Her Majesty's Stationary Office, 1998), which defines personal information and a series of principles that must be adhered to when organisations process any information that is deemed to be identifiable. These principles specify that identifiable information may only be held and used for the purposes it was collected for and should be destroyed once those purposes are complete. Equivalents can be found across Europe and are all conformant to the EU Data Protection Directive (European Parliament and Council, 1995). New Data Protection regulation that would apply directly in law to all EU members is also being drafted (European Commission, 2012a), which is discussed in more detail later in this in section.

In addition to the data protection legislation, the right to personal privacy is protected by the Human Rights Act (Her Majesty's Sationary Office (HMSO), 1998) and the Common Law Duty of Confidentiality mandates that information provided by an individual in confidence may only be shared more widely with the express consent that individual (Department of Health, 2003). This raises a significant issue for research projects processing hundreds of thousands of peoples' records: the information within them has been shared with a health service provider in confidence and for the purpose of direct care provision. However it is not usually

reasonably practicable to get informed consent from all participants for research uses of their data, particularly if there are hundreds of thousands of them.

Figure 3: Legal Framework

The Health and Social Care Act 2001 (Her Majesty's Sationary Office (HSMO), 2001) provided a legal basis for exemption from the Common Law Duty of Confidentiality to be granted by the Secretary of State for Health under Section 60 to manage such cases where it was not practicable to seek consent and the purpose of access could demonstrate a substantial public interest for that exemption. This was superseded by the National Health Service Act 2006 (Her Majesty's Stationary Office (HMSO), 2006) Section 251, which is discussed under section 3.1.2 below. Other legislation includes the Human Tissue Act 2004 (Her Majesty's Sationary Office (HMSO), 2004), was passed into law to make explicit special provisions regarding consent and participation where human tissue samples were concerned.

In 2012, a new Health and Social Care Act was passed by the UK Parliament for England and Wales, which established the Health and Social Care Information Centre (HSCIC) (Health and Social Care Information Centre, 2014a) as a body that could process "weakly pseudonymised" information (Her Majesty's Sationary

Office (HMSO), 2012). The Care Act of 2014 defined the Health Research Authority as the national body that would be responsible for overseeing research in the UK and to ensure that it complied with ethical review. IT also amended the Health and Social Care Act to enforce "…the need to respect and promote the privacy of recipients of health services and of adult social care in England…" with regards to the general duties of the HSCIC (Her Majesty's Sationary Office, 2014). The US Health Insurance Privacy and Accountability Act represents an example of international legislation that has sought to provide a basis in law for secondary uses of healthcare records.

To help interpret the legal framework in the UK by way of example, the HSCIC and NHS England both offer a series of guideline documents to help provide assistance in determining good practice and abiding by the law. Many of the more current documents were developed after the first Caldicott Review in 1997 (The Caldicott Committee, 1997), which was commissioned to help ease anxieties that were developing about the use of healthcare records. This review and the more recent Information Governance review of 2013 (Department of Health, 2013b) are discussed in detail in section 3.3. For example, two principles established by the 1997 review and reiterated by 2013 were the first, which stated that uses of identifiable information should be justified, and the second, which stated that identifiable information should be used when necessary.

In 2003, the Department of Health issued the Confidentiality Code of Practice (Department of Health, 2003), which was supplemented in 2010 to include details about public interest disclosures (Department of Health, 2010b): this made clear the requirements of the duty of confidentiality to clinicians and other secondary users, which are bound to uphold the same duty of confidence and needed guidance on what that entailed. This code emphasises the importance of using identifiable information only when absolutely necessary and encourages a culture of caution regardless of whether consent has been sought or the information is identifiable. An additional guide to confidentiality was released by the HSCIC in September 2013 (The Health and Social Care Information Centre, 2013) and these are discussed further in section 3.2.2.

An alternative interpretation of the law is that if records are rendered anonymous, the Data Protection Act does not apply to the records and there is no need to gain consent for reuse of the information in line with the Common Law Duty of Confidentiality. Definitions of what constitutes personal, identifiable and anonymous data are available in the DPA and from the independent public body called the Information Commissioner's Office (ICO) (Information Commissioner's Office, 2014b). The ICO offers key definitions from the DPA, including personal data (Information Commissioner's Office, 2014d), what constitutes personal data (Information Commissioner's Office, 2011b) and anonymisation guidelines (Information Commissioner's Office, 2014e). The ICO is responsible for ensuring compliance with the DPA and has powers to levy fines and proceed to prosecute organisations and individuals who are in breach of the provisions in the Act (Information Commissioner's Office, 2014f). The ICO also provides guidelines on how to share information, specifying Data Sharing Agreements, which detail how organisations agree to share information, including undertakings for how they intend to protect it as well as privacy impact assessments, which any organisation should run when it considers new processing of identifiable information.

The future of data protection legislation in the UK and Europe is unclear and contentious: the European Union announced plans to update the Data Protection Directive with a new Data Protection Regulation (European Commission, 2012a). In the UK, research bodies such as the Wellcome Trust have stated their position in partnership with a series of non-commercial, academic organisations, highlighting the difficulties and risks to continuing high quality health research if the proposed regulation is adopted in its current state (Wellcome Trust, 2014a). The literature is starting to assess the potential complexities of current and the new proposed European legislation on both care and research activities (Virone, 2012, Schutze, 2013), whilst Nyren et al. identified risks to the future of public health registers, threats to epidemiological research and that mandatory pseudonymisation (described in the next section) of healthcare records could render research databases "…useless for epidemiological research…" (Nyren et al., 2014).

3.1.1. Balancing Privacy Protection and Research Needs: Participant Anonymity and Risks of Identification

Rendering information anonymous and / or defining stringent access policies and disclosure controls are recognised as key solutions to protecting privacy and confidentiality (Agrawal and Johnson, 2007, Blobel, 2004, Carrion Senor et al., 2012). There are several examples of methods for rendering records anonymous: these include removing all primary identifiers from data sets (Information Commissioner's Office, 2014e) and aggregating results (Horner, 1998, Tamersoy et al., 2012). The literature is replete with concerns that anonymous data "are unlikely to be much use for research" (Academy of Medical Sciences, 2006) or have limited utility (El Emam and Dankar, 2008), and that complete anonymity is impossible (Sweeney, 2002, Ohno-Machado et al., 2001).

The linking of routinely collected records held within research and statistical data sets is critical for advancing research and building a more complete picture of individual and population outcomes. This is achieved by looking for matches of discriminating, (very often) identifying information or unique, national and organisational identifiers such as the NHS Number (Health and Social Care Information Centre, 2014c) between data sets. By attempting to link data sets with each other, a more complete profile is developed of an individual and they are therefore more likely to be identified (Thomas and Walport, 2008, Sethi and Laurie, 2013).

Several methods have been proposed to increase the utility of data (Tamersoy et al., 2012, Loukides and Gkoulalas-Divanis, 2012, Hughes et al., 2014, Ye and Chen, 2011). Pseudonymising records, whereby a link remains between the anonymous records that are released for research and the identifiable records, is also recognised as a means to preserve utility of de-identified records (Neubauer and Heurix, 2011, Aamot et al., 2013, Noumeir et al., 2007, De Lusignan, 2014) and permit record linkage, including various methods to mask or obfuscate data values (Kuzu et al., 2013, Ganslandt et al., 2011, Fernandes et al., 2013). For the purposes of linking data sets, encoding mechanisms such as Bloom Filters (Bloom, 1970) have been proposed whereby identifiable, discriminating data is encoded and

these encodings are used to match records across different data collections (Schnell et al., 2009). Unfortunately, these encodings are susceptible to cryptanalysis, meaning that the data could be read in the clear (Kuzu et al., 2011, Durham et al., 2013).

Sweeney observes that removal of highly identifying attributes is insufficient to prevent re-identification of individuals from the released data and offers k-anonymity based on statistical methods as a more robust system to assure a level of anonymity (Sweeney, 2002). Kalra and Ingram nevertheless emphasise the imperfections of anonymisation and pseudonymisation techniques and allude to the lack of "…consensus on good practice … in achieving pseudonymisation", which also emphasises that data may remain potentially identifying even after being de-identified (Kalra and Ingram, 2006), whilst Chalmers and Muir question the clarity of what levels of anonymisation would be acceptable in practice for patients (Chalmers and Muir, 2003). The literature also recognises that a profile may be established based upon information that has already been released to other parties (Li et al., 2011), prompting proposals for statistical disclosure control (SDC) (Martinez et al., 2013, Smith and Elliot, 2008, Elliot et al., 2006) and differential privacy algorithms (Mohammed et al., 2013).

Whether consent is sought, records are rendered anonymous or pseudonymous or exemptions are granted, there remains a risk to the individual that their records could be viewed by an unauthorised party and that they could be identified. Exeter et al. illustrate this with the privacy access continuum framework (Exeter et al., 2014), which incorporates geographical aspects to the potential for identification represented in Figure 4. Re-identification is clearly a risk regardless of the measures taken to maintain the anonymity of participants. It therefore falls to the researchers who are processing the information to treat it with due diligence. Their behaviour relies on effective training, education and good information security policy specification so that information is not shared inappropriately either by accident or deliberately, unauthorised access is prevented and risks to participants are minimised, particularly if they have not been asked for consent. Additionally, the need to ensure a clear legal and ethical framework is established when handling information even when anonymised is

emphasised (Toccaceli and Masocco, 2012) and a proportionate framework of information governance is encouraged (Sethi and Laurie, 2013).

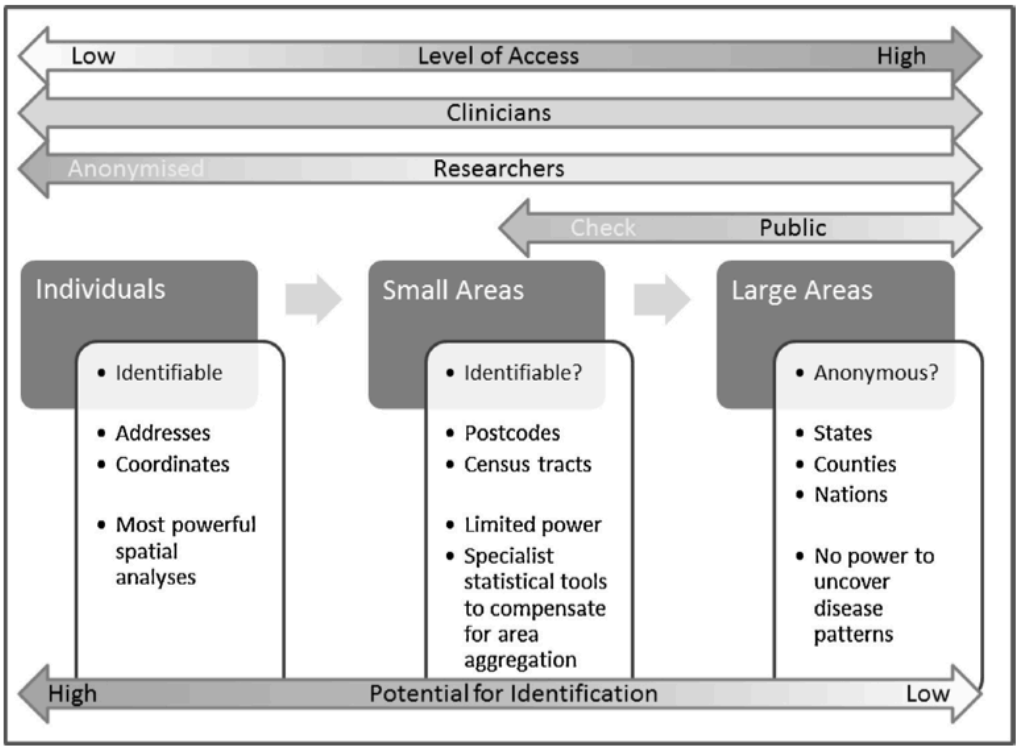


Figure 4: The privacy-access continuum framework (Exeter et al., 2014).

3.1.2. Identifiable Data Use and Participant Consent: Reasonableness and Practicalities

In recognition of the importance of sharing information and the use of identifiable or re-identifiable data outside of providing care, the UK enacted further legislation that would set aside the Common Law Duty of Confidentiality in cases where it is necessary to process identifiable health and social care records, it is not reasonably practicable to get consent from all the individuals about whom the information has been recorded and a substantial public interest can be established. This exemption was originally enacted in the Health and Social Care Act 2001 under section 60 (Her Majesty's Sationary Office (HSMO), 2001). This was later superseded by section 251 of the NHS Act 2006 (Her Majesty's Stationary Office (HMSO), 2006), which empowered a statutory board called the National Information Governance Board (NIGB) (National Information Governance Board (NIGB), 2013b) to consider applications for Section 251 exemption in their Ethics

and Confidentiality Committee (ECC) (National Information Governance Board (NIGB), 2013c) and make recommendations to the Secretary of State for Health that the common law of confidentiality be set aside (National Information Governance Board (NIGB), 2013a).

The Confidentiality Advisory Group (CAG) (National Health Service, 2014a) of the Health Research Authority (National Health Service, 2014b) took on the role of the NIGB ECC when the Health and Social Care Bill of 2012 was passed into law (Her Majesty's Sationary Office (HMSO), 2012). In order to be considered for Section 251 exemption, applicants must show a level of compliance with the Information Governance Toolkit (IGT) (Department of Health, 2010b) as well as demonstrate substantial public interest to be considered, the importance of which is recognised as being essential for maintaining public trust (De Lusignan, 2014). The IGT is discussed later in the chapter.

Individuals, however, prefer to be asked for consent when their medical records are used for research (King et al., 2012, Willison et al., 2003). In cases where consent must be sought, the UK Government has tried to simplify the process of gaining consent from individuals on a large scale and not on a case-by-case basis. Some of the recent centralisation projects including the Summary Care Record sought to apply an opt-out strategy (Watson, 2006, Health and Social Care Information Centre (HSCIC), 2014b) to comply with the law, whereby an individual would have to explicitly opt out of the schemes when they visited their General Practitioner (GP). This approach has also met with increasing unpopularity (described in section 3.3) and an opt-in approach, where people are explicitly invited to participate in and consent to research has been advocated (Ward et al., 2004).

There are concerns, particularly for research projects, that an opt-in approach adversely affects participation numbers (Trevena et al., 2006, Angus et al., 2003) and bias results of research projects (Al-Shahi et al., 2005, Junghans et al., 2005), whilst research ethics committees prefer an opt-in approach, the reasonableness and meaningfulness of consent has also been scrutinised (Williams et al., 2007, O'Neill, 2003, King et al., 2012, Corrigan, 2003). Other work indicates the need for both opt-in and opt-out (Langanke et al., 2011) as well as for transparency in the

use of healthcare information for research (Wiesenauer et al., 2012). The overhead of gathering consent for research purposes has been shown to adversely affect participant numbers in the case of clinical trial recruitment and research, specifically where the successive compliance mechanisms become increasingly burdensome (van Staa et al., 2014).

3.1.3. Managing Risks to the Rights of Participants: Regulation, Ethical Review and Permission to Conduct Research

Regardless of whether de-identified information is used, consent has been sought or exemptions have been granted, all uses of healthcare records are subject to ethical review, usually by a research ethics committee. Thompson remarks that research ethics committees have become an internationally recognised practice since the 1960s and '70s in response to unethical research practices (Thomson, 2012). In the case of research uses of data, until 2013 the National Research Ethics Service (NRES) (National Health Service, 2011a) that formed part of the National Patient Safety Agency (NPSA) provided research ethics committees (RECs).

NRES had a remit to "…to protect the rights, safety, dignity and well-being of research participants … and … to facilitate and promote ethical research that is of potential benefit to participants, science and society..." (National Health Service, 2011a). Where a use of data is classified as research, a research project must apply to a REC, the role of which is to "…review applications for research and give an opinion about the proposed participant involvement and whether the research is ethical…" independently of a project's funders (National Health Service, 2011b) where identifiable information is sought and whether consent has been granted by the subjects of the information or not.

The function of NRES has recently been taken over by the HRA, which provides guidance for each stage of the research process and "…works with (a range of bodies which have roles in regulating different aspects of health research in humans) to coordinate the overall system for regulation and governance of research…" in the UK (National Health Service, 2014d). The HRA provides access to the Integrated Research Application System (IRAS) (National Health Service,

2014c), which allows research projects to apply for permission and regulation by the relevant bodies.

NHS RECs are one such body and are made up of up to eighteen members including a portion of lay representatives (either not healthcare professionals or researchers). Like the NRES examples, they are responsible for "…safeguard(ing) the rights, safety, dignity and well-being of research participants, independently of research sponsors…" and will review applications for running research projects, clinical trials and establishing research databases amongst other research related endeavours that may affect the patient's clinical outcomes (National Health Service, 2014e). Research institutions such as UCL offer their own RECs, which may be approached instead of the NHS RECs if research does not involve the NHS (University College London, 2014c). There are also exemptions to needing ethical approval to perform research, listed under the UCL example (University College London, 2014b). Research councils also offer guidelines on conducting ethical research. The Medical Research Council for example offered a set of guidelines on good research practice as part of its ethics series (Medical Research Council, 2012).

These guidelines emphasise the importance of research excellence, transparency, honesty, openness and accountability bound to ethical principles. These are reflected in the data management plans that the MRC and other funders have started to develop, which require that research applicants specify how they plan to store and manage any data derived from electronic healthcare records, ensuring that "…research data are of the highest quality; have long-term validity; are well documented, so that other researchers can access, understand, use and add value to them over the decades and independently of the original investigators" and that "…research information about people is managed to the highest, most appropriate ethical and best standards…" (Medical Research Council, 2014a). The Wellcome Trust has similar guidelines and expectations that data sets are protected but also shareable for further research where permitted (Wellcome Trust, 2014b).

The literature has scrutinised the roles, constitution and conduct of RECs. Humphreys et al. illustrate the dominance of the medical profession on ethics committees in the UK (Humphreys et al., 2014). Savulescu et al. question the ethics

of RECs, claiming that they approved research that was unnecessary and under reported their approvals (Savulescu et al., 1996). Angell et al. and Al-Shahi et al. point to some inconsistency in the decisions made by different RECs (Angell et al., 2006, Al-Shahi Salman et al., 2014).

The literature has also identified cases where ethics committees have been deleterious to conducting research. Slowther et al. refer to the process of ethical review as being unethically bureaucratic and inflexible in decision making (Slowther et al., 2006), Ward et al. allude to the cumbersome process of gaining ethical approval and having to abide by different expectations from the range of ethics committees when seeking to recruit participants in a national case study in the UK (Ward et al., 2004) and Hemminki refers to the extra burden ethics committees put on clinical trials, specifying that their role should be "…advisory and they should not be censoring and preventing research, but advising and helping researchers to carry out responsible research…" (Hemminki, 2005). Boynton indicates that some ethics committees can be supportive and help guide good research, but this depends on the committee that reviews the application (Boynton, 2005).

This section has described the "why" and the "what" of information governance and discussed the literature around the effectiveness issues with these areas and approaches. It has included the legislation relating to privacy, confidentiality and data protection that is pertinent to the sharing of EHRs for care and for research purposes. It has provided an overview of the approaches and technologies that are used to maintain compliance and the organisations that are responsible for ensuring this compliance. It has also included discussions about de-identification and the risk management, regulation of research and ethical review process. The next section discusses the "who, where when and how" of information governance, which focuses on information security management and how it is used to meet the requirements outlined in the section.

3.2. Information Security: Practicalities and Guidance for Users

Figure 5 lists the standards and guideline documents that have been reviewed in this work. These standards provide a framework for the practical implementation of security management, privacy protection and data protection.

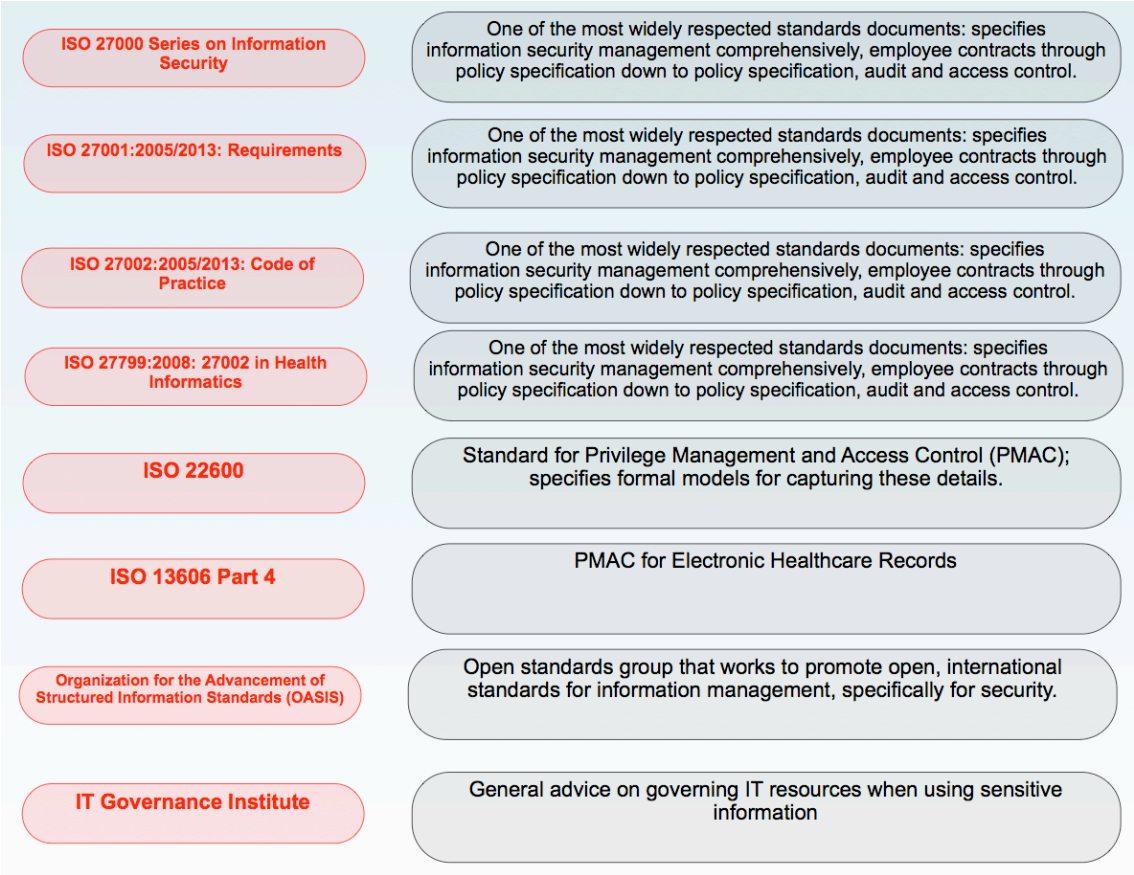


Figure 5: International Standards and Guidelines for Information Governance and Security

The ISO 27000 series covers general information security and is the focus of this subsection. The other standards are described in more detail throughout this and the following chapter, as their focus is more specific: for example, EN ISO 13606 part 4, which handles privacy management and access control (PMAC) is specific to the communication of electronic healthcare record extracts and is discussed in section 4.6.1, whilst the OASIS standards focus specifically on the technical implementation of access control and privilege management using scripting and mark-up languages, described in section 4.6.2. A series of different advisory sources exist for information security, for example some guidelines from industry, including the IT Governance Institute (ITGI) (Information Technology Governance

Institute, 2014) and the British Computer Society (British Computer Society, 2014).

3.2.1. International Organisation for Standardisation (ISO) 27000 Series of Standards for information Security Management

The basis for good security practice is established in the ISO 27000 Series of standards, which has been used as a guide to develop many of the guideline documents issued by the Department of Health regarding information security and governance discussed later in this section. The ISO 27000 series is composed of twenty three standards, with another twelve in development. It provides guidance for all organisations on how to protect their information assets, and focuses on the importance of having senior management support within an organisation as well as supporting people involved in the operations of an organisation. Originally a British Standard 7799, this was adopted by ISO with some amendments, and was evolved into three newer standards, ISO 17799, which was the information security code of practice, and ISO 27799 (The International Organisation for Standardization, 2008), which remains the standard for information security management in health when using ISO 17799 / 27002. 17799 became ISO 27002 (British Standards Institute, 2013a) within the 27000 series, and a new standard, ISO 27001, which handles requirements for information security (British Standards Institute, 2013b). These amendments happened in 2005, and there have since been updates to ISO 27001 and 27002 in 2013.

The ISO 27000 series is significant because it underpins key aspects of information governance management for healthcare records in care provision and research. ISO 27799 has informed the NHS Code of Practice on Information Security (Anderson, 2007) and the IGT is the Department of Health's interpretation of ISO 27002. Research organisations have opted to achieve ISO 27001 certification, whereby procedures and practice are independently audited and certified by ISO certified auditors once they have undergone appropriate training http://www.bsigroup.co.uk/en-GB/iso-27001-information-security/iso-27001-training-courses/iso-27001-lead-auditor/. The focus of this work has therefore been on ISO 27001, 27002 and 27799.

3.2.2. Implementing the 27000 Series: Information Security Management Systems, Codes of Practice and Policies

ISO 27001 focuses on the core requirements of managing information security. It recommends that an Information Security Management System (ISMS) be established that oversees the development of information security policies, runs risk assessments and develops risk mitigation plans. The standard specifies the need to manage supplier relationships, delivery of services and the security of physical infrastructure (including rooms, network connections and communications). The standard also emphasises the need for management commitment to information security, the development of an asset register, periodic review of policy and procedure and penalties for breaching security. It refers to the need to abide by local laws, to establish measures of the effectiveness of security controls, conduct internal audits and to spread awareness of security issues to organisation members. ISO 27002 offers guidance for organisations on how to implement their ISMS, policies and manage engagement with the people who are governed by those policies. ISO 27799 applies this general framework to the healthcare domain.

The IGT is an example implementation of ISO 27002 and 27799. It combines these with legislative requirements including the Data Protection Act and Human Tissue Act to provide a comprehensive list of requirements that organisations processing health and social care data must abide by. A list of requirements for secondary use organisations is listed in Figure 6. These requirements are general and incorporate the core elements of information governance. They emphasise the need for clear guidance, training and expertise in handling information and provide high-level guidance and general statements for adherence.

Req No	Description
Information Governance Management	
11-101	There is an adequate Information Governance Management Framework to support the current and evolving Information Governance agenda
11-105	There are approved and comprehensive Information Governance Policies with associated strategies and/or improvement plans
11-110	Formal contractual arrangements that include compliance with information governance requirements, are in place with all contractors and support organisations
11-111	Employment contracts which include compliance with information governance standards are in place for all individuals carrying out work on behalf of the organisation
11-112	Information Governance awareness and mandatory training procedures are in place and all staff are appropriately trained
Confidentiality and Data Protection Assurance	
11-200	The Information Governance agenda is supported by adequate confidentiality and data protection skills, knowledge and experience which meet the organisation's assessed needs
11-201	Staff are provided with clear guidance on keeping personal information secure and on respecting the confidentiality of service users
11-202	Personal information is only used in ways that do not directly contribute to the delivery of care services where there is a lawful basis to do so and objections to the disclosure of confidential personal information are appropriately respected
11-205	There are appropriate procedures for recognising and responding to individuals' requests for access to their personal data
11-206	There are appropriate confidentiality audit procedures to monitor access to confidential personal information
11-207	Where required, protocols governing the routine sharing of personal information have been agreed with other organisations
11-209	All person identifiable data processed outside of the UK complies with the Data Protection Act 1998 and Department of Health guidelines
11-210	All new processes, services, information systems, and other relevant information assets are developed and implemented in a secure and structured manner, and comply with IG security accreditation, information quality and confidentiality and data protection requirements
Information Security Assurance	
11-300	The Information Governance agenda is supported by adequate information security skills, knowledge and experience which meet the organisation's assessed needs
11-301	A formal information security risk assessment and management programme for key Information Assets has been documented, implemented and reviewed
11-302	There are documented information security incident / event reporting and management procedures that are accessible to all staff
11-305	Operating and application information systems (under the organisation's control) support appropriate access control functionality and documented and managed access rights are in place for all users of these systems
11-307	An effectively supported Senior Information Risk Owner takes ownership of the organisation's information risk policy and information risk management strategy
11-308	All transfers of hardcopy and digital person identifiable and sensitive information have been identified, mapped and risk assessed; technical and organisational measures adequately secure these transfers
11-309	Business continuity plans are up to date and tested for all critical information assets (data processing facilities, communications services and data) and service - specific measures are in place
11-310	Procedures are in place to prevent information processing being interrupted or disrupted through equipment failure, environmental hazard or human error
11-311	Information Assets with computer components are capable of the rapid detection, isolation and removal of malicious code and unauthorised mobile code
11-313	Policy and procedures are in place to ensure that Information Communication Technology (ICT) networks operate securely
11-314	Policy and procedures ensure that mobile computing and teleworking are secure
11-323	All information assets that hold, or are, personal data are protected by appropriate organisational and technical measures
11-324	The confidentiality of service user information is protected through use of pseudonymisation and anonymisation techniques where appropriate
Clinical Information Assurance	
11-400	The Information Governance agenda is supported by adequate information quality and records management skills, knowledge and experience
Corporate Information Assurance	
11-601	Documented and implemented procedures are in place for the effective management of corporate records
11-603	Documented and publicly available procedures are in place to ensure compliance with the Freedom of Information Act 2000
11-604	As part of the information lifecycle management strategy, an audit of corporate records has been undertaken

Figure 6: Information Governance Toolkit Requirements for Secondary Uses

After the recent information governance review (Department of Health, 2013b), the clinical lead of the review, Professor Martin Severs, stated at a conference on information governance that the IGT was not enough to provide a rigorous framework for protecting information. He advocated education rather than training, and user engagement, so that processors of healthcare records in any context fully understood what was required of them (Inside Government, 2013). The review is discussed in more detail in section 3.3.

In addition to the IGT, Figure 7 lists the other guidelines that have been reviewed as part of this work. These have been used to inform the requirements gathering process and development of the Secutype models described in Chapter 6. These guidelines were intended for managing EHRs and the security and governance requirements primarily in the care setting and their applicability remains for research uses, particularly in the case of identifying records. The guidelines below help to establish some of the specific details within the context of managing healthcare information and help people and organisations who process healthcare information to understand their responsibilities as specified in law and health service guarantees.

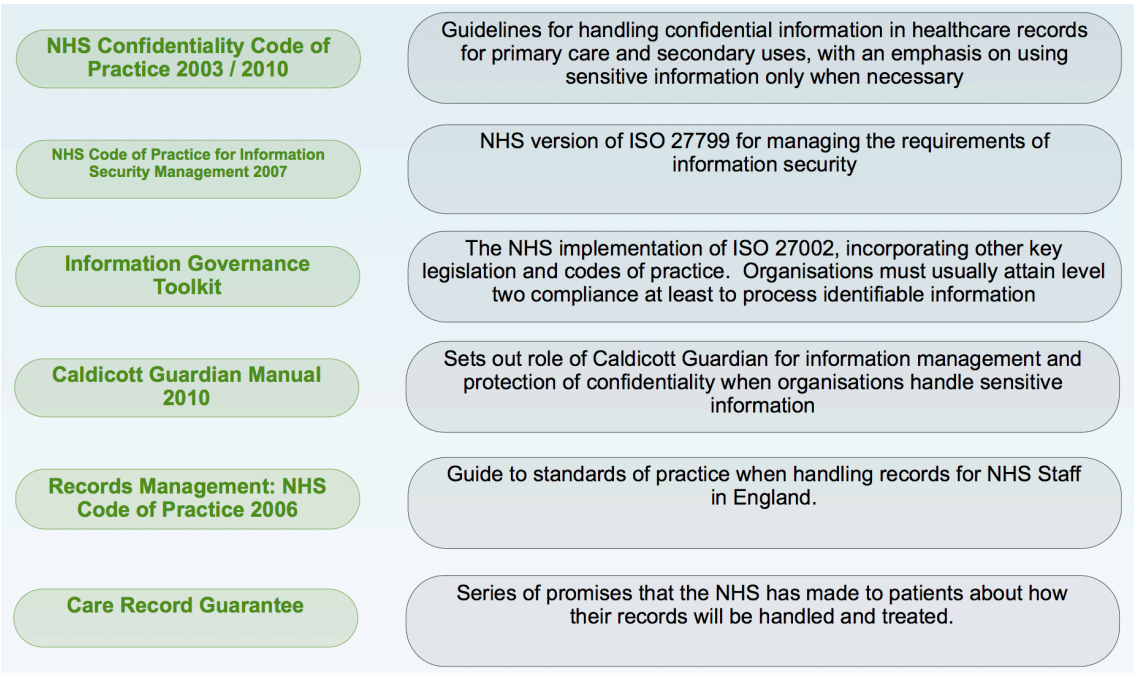


Figure 7: Codes of Practice and Guidelines Derived from UK Law and International Standards

The Confidentiality Code of practice has been discussed earlier in this chapter, along with the Information Governance Toolkit. The NHS Code of Practice for Information Security (Department of Health, 2007) summarises the requirements for managing information security, deriving the ISO 27799 standard for use within the context of NHS and associated organisations working practice. The Caldicott Manual (Department of Health, 2010a) provides a basis for the practice of Caldicott Guardians, who are recognised as the conscience of an NHS organisation when it comes to information sharing and appropriate uses in line with ethical and legal expectations. The NHS Records Management Code of Practice provides guidelines on how healthcare records should be handled and treated (Department of Health, 2006). The Care Record Guarantee outlines for the public how healthcare records about them and their treatment are handled and provides a series of pledges that the NHS makes about how they are protected. These guidelines have been designed to help inform the development of good working practice for handling healthcare records during care and secondary uses such as clinical research. This includes the development of information security policies for the variety of uses.

These guidelines are intended to help provide some assurance that policy specification and security reviews are complete and provide clear guidance on how to ensure that all the necessary information has been collected for a policy to be effective. The policy is the primary repository for knowledge in a given use setting for the management of information security, and its goals are to advise data handlers on how to behave with sensitive data resources. The governance policies must eventually be refined to computable heuristics for the technical security implementation. Creating, interpreting and implementing the policy correctly needs a thorough knowledge and understanding of not only the legal obligations in accordance with governance stipulations, but also the uses of the information, to where it will be sent, who will need to see it and what data they need to access. The policy must contain references to all other business relationships and third party responsibilities, as well as more technical, information technology oriented security policy specifications, which specify the configuration of technical solutions. The IGT and the other can advise at a general level on the development

of such policies, but reference must still be made to the original standards for a rigorous treatment of these requirements.

3.2.3. Issues with Policy Based Controls and Guidance

Issues with the use of policies have been identified. Of particular note are the points made by (de Lusignan et al., 2007) regarding the implementation of principles and policy into practice. This publication claims that the lack of definition of the finer data protection principles "leads to different interpretations and difficulties implementing the principles and policy into practice." They also find that "the degree of anonymisation of data that is necessary is unclear, as it is dependent on the type of data set, the purpose the data was collected for, the conditions under which the data was collected and the country." In addition, the publication states that the "[p]olicy needs to take account of individuals roles and complexity of the organisation. If the local policy is not specific, clear or directly relevant or if it does not address the specific tasks of the institution, the policy may be misinterpreted or necessary actions not taken. For example, the policy may need to address the security of paper files and emails in addition to electronic files" and asserts that "(e)nsuring that data protection policy is implemented is a significant undertaking that entails far more than writing a security manual." The paper concludes by suggesting a multifaceted approach to address "…the important organisational, technological, personnel and professional issues that impact on the implementation process…" and specifies the need for health informatics to "develop a core generalisable theory for implementing information governance policy that can be applied in everyday practice." The author submits that the proposal of this thesis of research to build a knowledge management framework represents a "core generalisable theory".

These views are supported by the recognition that very little work has been done on analysing the requirements for clinical research, or from a data release perspective (Manion et al., 2009). The statements by de Lusignan et al. are supported by the observations that have been discussed, as well as other commentary (Blobel, 2007). In addition to this, Laurie et al. have highlighted the varying interpretations of overlapping UK and European laws and that the Data

Protection Act 1998 Principles are "open to potentially varying interpretation in their implementation and operation in domestic systems…" and illustrates that this has meant that the Act is regarded as being overly restrictive for supporting clinical research (Laurie and Sethi, 2013). Becker discusses the risks of error in and complexities of policy interpretation (Becker, 2007), illustrating the issues of refining a policy that is human readable to one that is computable at a software run time level within several different software components. He discusses problems that ambiguity and incompleteness pose in a policy expressed in natural language.

These conclusions supported the identification of using a knowledge management approach, not unlike in the approach taken in health care data management where there have been a series of standards defining not only information models, but also models that would apply constraints as to how data could be represented for storage and retrieval as specified by the Clinical Archetype to form a generic EHR, which is described in section 4.4.3. There is no equivalent in the context of governance and information security controls. The benefits of using a knowledge management approach would be to help develop a common understanding and interpretation of the requirements for managing information governance, limiting the interpretation of generic guidelines and legal stipulations and allowing the modelling and definition of the ethical and legal principles, rights to privacy and duty of confidentiality in a way that can support information security good practice and the clear definition of policies to guide working practice.

3.3. Attitudes and Anxieties over Record Sharing and Challenges for Information Governance

Both the clinical care and research communities found that there were legislative, governance and good practice concerns about the handling of healthcare records and sharing for research, as discussed by Room (Room, 2004) and Walley (Walley, 2006). These concerns spawned a series of eSocial Science research workshops commissioned by the Oxford Internet Institute (Oxford Internet Institute, 2005) and (Oxford Internet Institute, 2005), where sociological researchers joined

biomedical and computer scientists to help understand the anxieties and consider the complexity of managing ethics concerns within the research context. In 2004, interpretation and implementation of governance stipulations and balancing individual rights to privacy against public good were discussed at a conference on the Governance of Medical Research Databases at the Royal College of Physicians (Royal College of Practitioners, 2004).

The discussion included the process of application to ethics committees and the confusion that ensued when these new eScience projects tried to get ethical approval. The conference occurred during an on going period of wider concerns about privacy invasion and the breach of human rights or data protection law, specifically with the establishment of centralised databases that are reviewed under the Joseph Rowntree Foundation commissioned report *The Database State*, which is one example that summarises many of these concerns (Anderson et al., 2009) and questioned the legality of these databases. Recently more evidence of data breaches has been published. The Board of the HSCIC commissioned a report by Sir Nick Partridge to review the data releases by the former NHS Information Centre, where amongst other findings, "…the review discovered lapses in the strict arrangements that were supposed to be in place to ensure that people's personal data would never be used improperly…" and that information which had been anonymised and released were rendered identifying after they had been linked with other datasets (Partridge, 2014).

3.3.1. Public and Professional Concerns

The literature observes that there are patient and research participant privacy concerns about healthcare information being shared more widely. Haga et al. found that research participants' anxieties about privacy breaches warranted a clear description of research project's data sharing plans (Haga and O'Daniel, 2011). Other studies have found that whilst there is an ambiguity about the understanding of confidentiality between different studies, there is broad participant support for data sharing to improve healthcare outcomes, however these are tempered by privacy and confidentiality concerns and the support is linked to the goals of research and perceived benefits of the outcomes (Jamal et al.,

2013). Van Staa et al. note that GPs taking the time to explain the benefits of research to their patients have a positive outcomes for participation, but that in areas of higher depravation, patients are likely to be "…too busy dealing with life stresses to be interested in research…" and recruitment might be affected in these areas (van Staa et al., 2014).

King et al. demonstrate support from patients, though also note at they expect to be asked for permission to use their healthcare records, want to know details about the research and have concerns that their information might be linked to more sensitive records (including sexually transmitted diseases and abortion registers) or be used by organisations outside of the clinical care provision and research settings (King et al., 2012). This is echoed by Perera et al., who show that there is broad support for computerisation of healthcare records to support care and secondary uses whilst recognising that there are privacy risks, though there is a substantial minority who have privacy concerns over sharing data for secondary purposes (Perera et al., 2011). Caine et al. highlight that patients want granular control over the privacy of their EHRs (Caine and Hanania, 2013) and Callway describes an IPSOS MORI poll supported by the Wellcome Trust, which shows that sixty per cent of 1,396 UK adults would be willing to participate in "…medical research project which involved allowing access to (their) personal health information (medical records)…" (Callway, 2013).

Anxieties have nevertheless remained and arguably increased: the updates to the Health and Social Care Act 2012 that were underway in 2011 caused a series of information governance concerns given the proposals for wider sharing with industry, and this is highlighted by the NHS Future Forum report on information, which called for a review of information governance to support the sharing of information as part of the government's vision of the NHS and general acknowledgement of the importance of the sharing of information to support healthcare (The NHS Future Forum, 2012). This prompted the second Caldicott review of information governance (Department of Health, 2013b) at the behest of the then Secretary of State for Health Andrew Landsley chaired by Dame Fiona Caldicott, who led the first information governance review in 1997 in the wake of

concerns regarding centralisation of patients' records (The Caldicott Committee, 1997).

The information governance review acknowledged the importance of sharing information by developing an additional principle, that it is sometimes as important to share information as it is to protect confidentiality. It makes further recommendations about how information should be processed, developing the data safe haven concept to include accreditation and independent certification described later in the chapter. The most recent example of anxieties has been the outcry over Care.data: this initiative, run from the HSCIC, is seeking to "fill in the gaps" of healthcare records and link them to social care and genomics data (Department of Health, 2013a).

Public and professional anxieties over how this information would be processed and to whom it would be made available were clear: a petition was launched by pressure group 38 degrees, where one hundred and twenty thousand petitioners sought to opt out of the scheme, helping to bring about a six month delay to the project (Band, 2014). The Royal College of General Practitioners issued their own statement that, whilst still broadly supportive of the initiative, specified their concerns over it (Royal College of General Practitioners, 2014) and poor public engagement and referred to their own guidance on secondary uses of electronic healthcare records (Royal College of General Practitioners, 2013). A Nature editorial also highlighted the poor handling of the public engagement (Nature Editorial, 2014a), and a further editorial describes how the right for people to opt out of the initiative is being "downplayed," the engagement with the public is not optimal and that the removal of primary identifiers is not a fool-proof solution and risks to the participants remain (Nature Editorial, 2014b). This has resulted in a six-month delay to project whilst the UK Government engages further with the public to explain how privacy and confidentiality will be protected. Sheather and Brennan also highlight some of the risks involved and allude to low public confidence in the government's ability to manage personal data (Sheather and Brannan, 2013). Rynning emphasises the importance of public trust for the success of EHR systems (Rynning, 2007) and Anderson refers to the privacy concerns of both patients and practitioners acting as a barrier to the adoption of

electronic healthcare records, "…since many EMR systems are Web-based, many physicians and patients fear that medical records may not be secure..." (Anderson, 2007). Anderson has also referred more generally to the challenges of "…provide(ing) the data required by the new forms of health care delivery and at the same time protect the personal privacy of patients…" (Anderson, 2000).

3.3.2. Legal Enforcement

There are some examples of prosecutions as a result of breaking the law when handling EHRs, all of which relate to the context of care provision. The Information Commissioner's Office maintains a publically available website that includes details of enforcement, warnings, fines and prosecutions for breaching the Data Protection Act in the UK (Information Commissioner's Office, 2014f). One example of precedent is the EHCR ruling in favour of P against Finland for a claim of data misuse because the respondent could not provide evidence that they were protecting the data as they claimed (HealthImaging, 2008); this highlights the importance of maintaining a record of how individuals' data items are protected both at the point of collection and thereafter, and the importance of maintaining an effective audit.

3.4. Summary of Expectations and Issues

This chapter has provided an overview of the principles of information governance and focused on the importance of participant anonymity when reusing records for research with or without consent. Information governance involves legally protected rights to personal privacy, data protection and medical professional duties of confidentiality and ethical review, all of which are upheld by practical implementation of security practices to safeguard these legal and ethical requirements. International standards form the basis for organisational guidelines through a series of government funded and independent organisations and local implementation.

The basis for information governance is subject to frequent change and update: the future of European data protection legislation is unclear given the proposed changes to the Data Protection regulation at the European Commission and the ISO

Standards themselves have undergone a substantial review and update between 2005 and 2013. Additionally, the process of ethical review has remained under scrutiny both in terms of its consistency and outcomes and whilst it appears that public good is favoured over individual rights to personal privacy there is a general disagreement over whether this should be the case.

The guidelines and standards, which are regarded as the foundation of good practice, are themselves regarded as sometimes unclear, unhelpful and not easily implemented. Whilst the sharing of healthcare information is generally supported, the public remains anxious and sceptical about the protection of their records and some uses, and prefer to be kept informed about how research is using the records. Regardless of the legislative exemptions to gaining explicit consent, reassurance and engagement are recognised in the literature and wider commentary. It is clear that there have been failures in public engagement for some of the more high profile secondary use and collection projects in the UK, specifically Care.data.

It is however clear that participant consent is regarded with high importance, though particularly in the UK, having people proactively opt-in or opt out of research participation has remained an unanswered yet hotly debated area of contention. The reliance on using de-identified records as a means to satisfy the Common Law Duty of Confidence and data protection requirements where participant consent cannot reasonably be sought is one that is repeatedly tested: there is still a risk of re-identifying individuals regardless of the methods and practices applied to processing participant records to maintain their anonymity. A compelling practice of linking records to provide a clearer picture of healthcare outcomes when used for research increases the risk of re-identifying individuals, and methods to enhance participant privacy have been shown to be susceptible to attack.

It is very clear from the literature that the process of applying information governance requirements needs support, simplification and clarity for people and systems that process healthcare records when used for care and research. Public trust is also essential for not only ensuring that people are content to participate in research, but they are also aware that records may be reused without their knowledge in the current climate. Given that there are always risks to people

about whom the records are being kept, reassurance that information is being processed appropriately and safely and procedures for this are understood and being followed has continued to become more critical, particularly if the public in the UK and beyond are going to continue to support reuse of health and social care information.

Chapter 4. Developments in Healthcare Information Strategy

The literature, legislators and society expect that a substantial set of legislative, working practice guidelines and information security procedures to go some way to mitigating the risks involved with using people's records for care and research. It is however clear that some risks to patients and their privacy, their trust in healthcare services and the medical profession's ability to discharge their duty of confidentiality remain. Despite increasing public anxiety about the use and sharing of personally identifiable information, these risks appear to be accepted by governments across Europe and beyond, balanced as they are with the care, research and wider benefits of sharing healthcare information and some public support for research use. Government support is clear from the financial and policy support to implement IT systems in healthcare settings across the world, which has continued for several decades, to support accurate and timely sharing through the development of electronic healthcare record systems.

For example in the UK, the first attempts to implement networked, EHR systems have been in progress since the 1990s, as discussed by Kalra (Kalra, 2002). The NHS has historically invested in such systems, and at times directed the market and evolution of these systems through national information management and information technology strategies, as summarised by the NHS in 2001 in its report *Building the Information Core – Implementing the NHS Plan* (Department of Health, 2001), which emphasised the development of network and other IT infrastructure, electronic records and both nationwide and local applications. These successive strategies and investments have helped to bring some of the benefits of health care data sharing to fruition.

The most recent and by far the largest strategy for IT provision within the NHS has been the UK NHS National Program for IT (NPfIT), implemented by the Connecting for Health (CfH) initiative. Considerable investment has therefore recently been made in infrastructure, including the establishment of the NHS Net N3, and increased provision of resources for Information Technology and

Management (IM&T) departments across various healthcare trusts, so that they are better equipped with IT resources, hardware, software and personnel support. CfH originally had a remit to "…maintain and develop the national IT infrastructure…" (Department of Health, 2011).

CfH exemplified the reuse of health care data for purposes other than the provision of care, specifically research, commissioning of services and reporting activities, supported under the Secondary Uses Service (Health and Social Care Information Centre, 2014d). CfH was disbanded in 2013 as a result of a National Audit Office review of its effectiveness and reviews of ongoing information processing requirements (National Audit Office, 2011), which meant that many of the functions of CfH were taken over by the Health and Social Care Information Centre (HSCIC) in 2013, following provisions made within the Health and Social Care Act of 2012 (Her Majesty's Sationary Office (HMSO), 2012). These provisions included more support for the reuse of healthcare records to support clinical research, as well as commissioning, clinical trials recruitment and public health surveillance.

In 2006 The Chancellor of the Exchequer and Secretaries of State for the then Department of Trade and Industry, now the Department for Business, Innovation and Skills (Department for Business, 2014), asked Sir David Cooksey to review and independently advise "…on the best design and institutional arrangements for the public funding of health research in the UK…" This review identified that there were many strengths to the UK Health Research System, with a tradition of "…producing excellent basic science…" and citing the UK Medical Research Council (MRC) funding twenty-seven Nobel Prize winners since 1913. The Review also "…found … that the UK is at risk of failing to reap the full economic, health and social benefits that the UK's public investment in health research should generate…" due to no overarching UK research strategy, a gap in "…translating ideas from basic and clinical research into the development of new products and approaches to treatment of disease and illness; and implementing those new products and approaches into clinical practice…" Amongst other recommendations, it emphasised the importance of allowing research access to the records that were to have been stored within the CfH Spine and that the MRC

should build on its links with other Research Councils. It also recommended that the National Institute for Health Research (NIHR) be established as "...real, rather than virtual institute, established as an executive agency of the Department of Health…" (Cooksey, 2006).

The 2008 Data Sharing Review by Walport and Thomas also emphasised the importance of sharing information for research purposes, but was performed at the request of the Prime Minister in his Liberty Speech of 2007. The goal of the Review was to "…review of the framework for the use of personal information in the public and private sectors…" and consider any changes needed to the operation of the Data Protection Act 1998. The result of the review was a series of recommendations that were designed to "…transform the *culture* that influences how personal information is viewed and handled; to clarify and simplify the *legal framework* governing data sharing; to enhance the effectiveness of the *regulatory body* that polices data sharing; to assist important work in the field of *research* and statistical analysis; and to help safeguard and protect personal information held in publicly available sources…" (Thomas and Walport, 2008).

Three key recommendations were applicable to research: Recommendation 15 was to develop a sae working environment for researchers called *safe havens*, which are described in section 4.3; Recommendation 16 was that Government departments and others hoping to develop datasets for research and statistical purposes to work with academic partners to set up these safe havens; and Recommendation 17 was that the NHS should "…develop a system to allow approved researchers to work with healthcare providers to identify potential patients, who may then be approached to take part in clinical studies for which consent is needed…"

The research community in the UK and Europe have therefore, in parallel with the care service reform and strategies, made significant provision to support the use and sharing of healthcare data and develop a means to address clinical research questions. This chapter discusses the strategy and infrastructure that have been used to develop these clinical care and research systems and services both in the UK and beyond. It describes the clinical research strategy adopted by UK and other EU governments to support clinical research. It introduces the

international electronic healthcare record standards, which provide the basis for the knowledge management for the correct, timely and accurate sharing of healthcare records and inspired and informed the development of the information governance knowledge model developed in this thesis and described in Chapter 6.

4.1. Fostering Collaboration for Clinical Research

The support for sharing and collaborating in research endeavours within the healthcare industry has been clear from funders and government commissioned independent reviews that have been described. A key example was the former Department of Trade and Industry funded UK eScience Initiative (The eScience Initiative, 2014), which was established to research the use of experimental, high performance computing resources to solve research problems across all academic disciplines. It was intended that not only a unified, 'grid' system of research data be built to share resources and findings, but also that researchers from different disciplines could share expertise and techniques to see whether they might support each others' work.

The MRC funded several eScience projects that worked with computing researchers to try to establish repositories of information that would help solve particular problems: eDiamond researched the use of high performance computing and network resources to distribute the workload of breast cancer screening (Oxford University, 2002); CancerGrid (Oxford University, 2008) looked at methods to store and refine data using the Web Ontology Language (OWL) (Worldwide Web Consortium, 2013); PsyGrid constructed a repository for remote data collection and integration with other data systems, which aimed to "…ascertain and characterise a large, representative sample of schizophrenics…" (Ainsworth and Harper, 2007); the Clinical eScience Framework (CLEF) built a repository of shareable clinical information, and researched the security and legal implications of this repository, how to enrich the data repository with data embedded in the free text of clinic letters and report narrative that was hard to query and had to be mined using natural language (Kalra, 2006) processing techniques, so that the data would be stored in a structured form and more readily queried (Rogers et al., 2004).

Biobanking has become another significant development in research and preventative medicine in the UK and internationally, where research repositories have been established to provide a common resource for researchers. An example includes the UK research councils and Department of Health supported UK BioBank. This initiative was developed to collect biological samples and other information about half a million participants to help provide resources for scientists to work on the prevention of a variety of conditions, including heart disease, depression and dementia amongst others (UK Biobank, 2011). Other examples around the UK exist, including facilities at UCL (University College London, 2014a), Oxford (Oxford Biobank, 2012) and the National Institute for Health Research (NIHR) (National Institute for Health Research, 2014b) biobanks. Condition specific biobanks have also been established: these include the Cancer BioBank (King's Health Partners, 2014) and the ME-CFS BioBank (London School of Hygiene and Tropical Medicine, 2014). Examples of international collaboration include an initiative for Jordan and neighbouring countries for biobanking, as described by Barr et al. (Barr et al., 2014).

In the UK, several initiatives are being developed to improve the flow of clinical data from hospitals and GP practices through to research projects. One initiative has been set up by the Farr Initiative funders to develop linkage methodologies, develop best practice across datasets and engage with the public: this is called the Health Informatics Research Network and is managed by the Farr Institute (Medical Research Council, 2014b), described in 4.2. Another example are the NIHR funded Biomedical Research Centres (BRCs): this initiative, which includes amongst others collaborations between UCL and UCL Hospitals NHS Foundation Trust that aim to conduct research on acute care, cardiometabolic, cancer, infection, immunity and inflammation and neuroscience (Research, 2014), South London and Maudsley NHS Foundation Trust that aim to work on translational research in dementia and other mental health conditions (Trust, 2014), and Oxford University hospitals, which are performing translational research across all therapy areas (National Institute for Health Research, 2014a). These initiatives rely on the sharing of information and are aiming to link their data across the BRCs, as well as to "…develop, design and provide common infrastructure,

standards and services, which will allow research users to fully exploit care data for research benefit…" (National Institute for Health Research, 2013).

Other examples of the collection of clinical information include the data platforms that serve national scales. The Scottish Health Informatics Programme (SHIP), funded by the MRC, Economic and Social Research Council (ESRC) and Wellcome Trust funded this programme to develop a Scotland-wide research platform to support the "…collation, management, dissemination and analysis…" of EHRs, building upon Scotland's history of nationwide record linkage research (Scottish Health Informatics Programme, 2014, Williams et al., 2010, de Lusignan et al., 2010). The Secure Anonymised Information Linkage (SAIL) Databank was an equivalent Welsh resource held at the University of Swansea, which brought together "…the widest possible array of routinely-collected data for research, development and evaluation…" in a "…Wales-wide research resource focused on improving health, well-being and services…" (Swansea University, 2013, Lyons et al., 2012). Both SHIP and SAIL developed their own information governance frameworks for securely working with EHRs, mandating a proportionate and effective approach to managing information governance (Jones et al., Sethi and Laurie, 2013). Both SHIP and SAIL have been continued within the Farr Institute for Health Informatics Research described in the next section.

In England, joint funding form the NIHR and the Medicines and Healthcare products Regulations Agency (MHRA) have established the Clinical Practice Research Datalink (CPRD) as "…the new English NHS observational data and interventional research service…" This service provides access to data sets derived from EHRs. It links data sets to the Office for National Statistics Mortality data and to census data (Clinical Practice Research Datalink, 2014). The Health and Social Care Information Centre runs a clinical data warehouse called Hospital Episode Statistics (HES) that processes "…over 125 million admitted patient, outpatient and accident and emergency records each year…" from hospitals in England. This resource has been made available to researchers and others, including patients and service providers, to "… reveal health trends over time…" and to "…monitor trends and patterns in hospital activity…" (Health and Social Care Information

Centre, 2014b). HES has also been linked to ONS mortality data and research projects have performed record linkage to HES.

4.2. Epidemiological and Personalised Medicine Strategy and Support

After the establishment of a series of larger scale data repositories, the MRC and other funders recognised that the importance of having robust, secure and extensive holdings. A recent example is a consortium of ten funders provided a £17.5M research award to fund research at the new Farr Institute of Health Informatics Research, which was further supported by a government commissioned £20M capital fund from the MRC to develop four centres of across the UK. This award has helped to develop infrastructure to support the linkage of larger data sets, which is anticipated to support the expansion and development of research queries, merging both health and social care records (Farr Institute of Health Informatics Research, 2014).

This funding has been invested to develop *Big Data* resources, where large amounts of information are processed by high performance computing and networks. The MRC and other funders, such as the ESRC have both supported the expansion of resources to help manage the "…enormous volume of data that is being collected by government departments, businesses and other organisations within the UK which can be used to the mutual benefit of academic research, organisations and society as a whole…" (Economic and Social Research Council, 2014b). These facilities are designed primarily to support larger scale epidemiological studies, using the process of record linkage to match health record information to social care records through the collaboration between Farr Institute and Administrative Data Research Network (ADRN) (Economic and Social Research Council, 2014a).

The higher capacity computing and data management infrastructures have also been essential for the advancement of genetic information processing: the 100,000 Genomes Project run by Genomics England emphasises the reduced cost and time taken to map individual people's genetic structure; the project aims to link the genetic information with patients' EHR data in order to "…help understand

diseases and to tease apart the complex relationship between our genes, what happens to us in our lives and illness…" (Genomics England, 2013). A similar initiative in the US exists with the Electronic Medical Records and Genomics (eMERGE) Network in the United States, where Gottesman et al. report on "…validating the concept that clinical data derived from electronic medical records can be used successfully for genomic research…" (Gottesman et al., 2013).

The expansion of Big Data facilities does not focus solely on epidemiological studies: genomics research is by definition specific to individuals and is recognised as significant for personalised health, particularly in terms of discovering trends across genetic profiles and benefits of treatments and "…a more tailored approach to prescribing…" (Haycox et al., 2014). Caulfield and Zarzeczny highlight the balance between defining medical need of personalised medicine and the challenges to the clinician's ethical and legal obligations that it represents using Canada as an example (Caulfield and Zarzeczny, 2014). Other initiatives continue to focus on individuals, including the *p-medicine* initiative, which has been developing an infrastructure to help "… bridge the gap between treatment given to patients and research to find better treatment for patients…" where "…the main drivers for such an infrastructure are clinicians as they have direct contact with patients…" and uses computer modelling and detailed healthcare records to develop a *virtual physiological human* to meet its goals (p-medicine, 2014).

p-medicine is an example of a European collaborative initiative, funded by the European Commission Seventh Framework Programme for Research and Technological Development (FP7) (European Commission, 2012b). Other examples of projects funded through the EC include the TRANSFoRm project, which aimed to develop a "… rapid learning healthcare system…" to support clinical research (TRANSFoRm Project, 2014) using information derived from care settings, which in turn can be used to improve care provision. The Electronic Healthcare Records for Clinical Research (EHR4CR) and European Medical Information Framework (EMIF) projects are also examples of research projects funded through the EC. EHR4CR is developing a series of tools that will support clinical trials recruitment across participating hospitals throughout Europe and this tooling will use information stored in patients' EHRs to determine their

eligibility for participating a trial, as well as manage their recruitment (EHR4CR Consortium, 2014). EMIF "…aims to create an environment that allows for efficient re-use of existing health data…" using a common information framework to support research in Alzheimer's disease and metabolic complications of obesity in the first instance across Europe (European Medical Information Framework (EMIF) Consortium, 2014). Both these projects are co-funded by the Innovative Medicine Initiative (IMI), which "…supports collaborative research projects and builds networks of industrial and academic experts… to boost pharmaceutical innovation in Europe…" (Innovative Medicines Initiative, 2014). EMIF and EHR4CR are discussed in more detail in section 5.4.

There are no authoritative exclusion criteria for whether a project or initiative is defined as a Big Data project or not: across the examples cited in this section and the previous section, a consistent theme has emerged between not only requiring safe use and appropriate governance of healthcare information, but also a means by which standardising the records can be achieved for a consistent understanding and data management strategy to support research. Clinical information sharing between the clinical care and research communities relies upon the adoption of common representations of the healthcare records and must occur in accordance with the information governance requirements described in Chapter 3.

4.3. Data Safe Havens

Data Safe Havens have been alluded to in previous chapters. The original intention for these in the research context was proposed in the Data Sharing Review (Thomas and Walport, 2008), which recommended that they "…should be developed as an environment for population-based research…in which the risk of identifying individuals is minimised…" They also recommended that researchers working in these safe havens be accredited and "…bound by a strict code, preventing disclosure of any personally identifying information, and providing criminal sanctions in case of breach of confidentiality."

This paradigm has been developed further by the Health and Social Care Information Centre as accredited safe havens, developed to handle commissioning services that use "weakly pseudonymised information" where a postcode and NHS

Number are retained. The accreditation requires IG Toolkit level two compliance, signing a statement of principles specified in a data sharing contract with the HSCIC and submit to an annual internal audit as part of the IG Toolkit level 2 compliance (Health and Social Care Information Centre HSCIC, 2013).

The Information Governance Review made further recommendations about the development of an accredited data safe haven and its applicability beyond commissioning. It recommended that a safe haven achieve certification from an independent body such as an ISO 27001 certification body, and have periodic, independent audits. It also recommended that the linkage of personal, confidential or de-identified data, given that linkage could increase the likelihood of re-identification, for any purpose other than direct care should occur within these accredited safe havens (Department of Health, 2013b).

4.4. Knowledge Modelling and Health Informatics: Standardised Electronic Healthcare Records and Modelling Paradigms

The capture, storage and sharing of information stored in electronic healthcare records has been supported by a series of standardised knowledge management architectures. Much of the software infrastructure that supported the implementation of CfH continues to handle healthcare records throughout the healthcare providers and in some cases supports clinical research, which have a focus on working with common standards for healthcare information management. A significant example of such standards are those relating to electronic healthcare records and are implemented using a combination of business level programming and database technology, which form the architecture for a knowledge management framework. The technologies used provide a tiered architecture for software (described in section 4.5), where users are presented with information, which they can review or amend. The previous section on research strategy shows a focus on the development and use of common standards for storing healthcare information. The approach taken to developing the international EHR standards have been an important basis for the development of the knowledge models designed in this thesis because they provide the basis for

the approach and design of the information governance knowledge models described in Chapter 6.

The clinical community recognised the importance of having an agreed structure for information that needs to be shared between the different healthcare professionals who would need to access those records and that a "silo mentality" for storing records was not in keeping with effective healthcare delivery and that a means to achieve a common understanding between practitioners and semantic interoperability between record keeping systems was key to supporting shared care (Kalra, 2002). The EHR was proposed as a means to achieve this, recognising the need that they be "…scalable, portable, distributed, and interoperable which has to be enabled by a proper architecture supporting informational and functional needs…" (Blobel, 2002) and follow "…advanced architectural paradigms…" (Blobel, 2006) or an "…integrated approach to formally modelled system architectures…" (Bernal et al., 2012). This is achieved by modelling clinical concepts according to an agreed structure between domain experts (Beale, 2002) and defining constraints on how information should be represented in order to achieve a consistent semantic interoperability. This section describes a series of standards and the engineering specifications that provide a basis for common understanding of clinical information and consistent use across a series of specialists and purposes. It discusses the technologies used to implement knowledge-managed services and concludes with a discussion of the protection mechanisms developed for electronic healthcare records.

4.4.1. EHR Standards

The EHR's primary purpose is to model reusable components so that data contained within the records can be shared as information consistently, which has both inspired and derived benefit from ongoing research into the application of EHRs in various clinical contexts (CHIME EHR Group, 2014, Goossen, 2008, Duftschmid et al., 2013, Kalra and Fernando, 2013). A community of international standards development and research groups have been working over many years to capture these requirements and paradigms in an attempt to garner international agreement and collaboration on how clinical domain knowledge should be

modelled and represented in record systems. Knowledge management has relied on the development of a series of standardised approaches that provide a consistent, standardised view of clinical concepts shared between clinicians who use them to treat patients, where there are examples of discharge summaries, diabetes and cardiovascular care, amongst others (Farfan Sedano et al., 2011, Moner et al., 2008, Austin et al., 2009).

The literature identifies this standardised approach as essential to achieving a common understanding between clinical users (Duftschmid et al., 2010) and illustrates the distinction between information and knowledge, where these approaches focus on modelling information and modelling how these information items relate to each other (Martinez-Costa et al., 2010). Examples of international standards and specifications for EHRs are described by Kalra and Goossen et al. (Goossen et al., 2010, Kalra, 2006), the most frequently occurring examples in the literature include those published by the openEHR Foundation, European Committee for Standardization (CEN), the International Organisation for Standardization (ISO) (The International Organisation for Standardization, 2011) and Health Level 7 (HL7).

openEHR is a "…virtual community working on the interoperability and computability in eHealth…" with a focus on the development of open standards for EHR models (openEHR, 2014b). It has developed a series of model specifications that include information models, which define data types (string values, numerical values and dates, for example.) and a "blueprint" for the structure of components of a record called an Archetype model, described in the next section. openEHR provides a set of Archetypes that can be obtained and used under open source license.

HL7 is a not for profit international organization that has operated since 1987 and "…provide(s) a framework (and related standards) for the exchange, integration, sharing, and retrieval of electronic health information. These standards define how information is packaged and communicated from one party to another, setting the language, structure and data types required for seamless integration between systems…" (Health Level 7 (HL7), 2014a). It does this in a similar way to openEHR, by defining a series of models for a Clinical Document

Architecture (CDA), Reference Information Model (RIM) for message exchange, and Templates, which provide a constraint model to the RIM model, much like the Archetype from OpenEHR.

CEN is a European standardisation agency, which focuses on a wide range of areas as well as healthcare (Standardization, 2014). The most pertinent example for EHRs is the EN 13606 standard, which is a five part EHR communication standard that defines a "dual model architecture," including a reference model that defines the basic entities for representing any part of an EHR, an archetype model that defines the basis for specifying formal models of clinical concepts, as well as a communication model, interface specification and privilege management and access control for these extracts (EN 13606 Association, 2014). The various parts of CEN 13606 were ratified as an international standard by ISO from 2008. This work focuses on the EN / ISO 13606 reference model, using it to implement the proposed knowledge management tooling described in Chapter 7.

There is a consistency between the overall architectures of the models that these standards have developed, which include a set for representing information items and a set designed to define a set of constraints on how these information items should be structured to represent clinical knowledge and concepts. The most important common feature of these EHR and clinical document models is that they specify a representation of properties that are common to clinical information in general (for example, a headed section within a clinical document) without defining the particular information structures and semantics of actual clinical data (such as a headed section for Past Medical history). The means of defining these particular information structures is explained in sections 4.4.2 and 4.4.3 below. Figure 8 provides an overview of the components within an EHR system, including those that are described in the forthcoming sections. The next section describes these elements, and how they are represented as reference or information models.

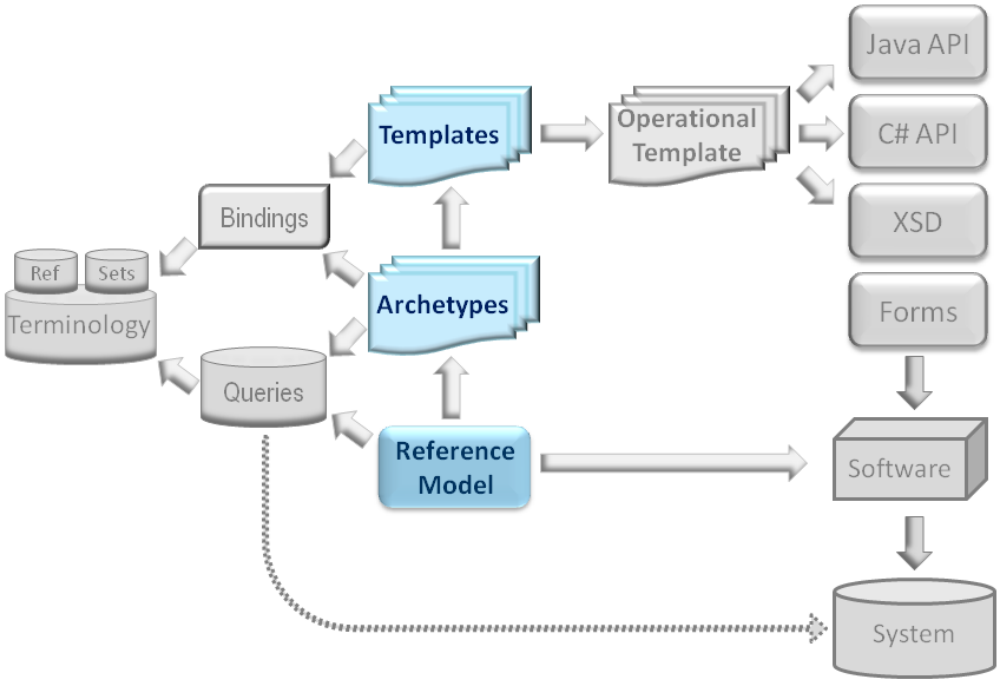


Figure 8: Component view of an EHR system from openEHR (openEHR, 2014h)

4.4.2. Information and Reference Models

Any information recorded at the point of care is subject to grouping and representation that must be clinically meaningful and compliant with medico-legal requirements (Kalra, 2002). The EN ISO 13606 standard provides a Reference Model to represent items of medico-legal relevance, as shown in Figure 9. There are a series of classes that group individual information items into a substructure of Record Components from EN ISO 13606 respectively. Table 1 from EN ISO 13606 part 1 describes these record components and their purpose. The intention of these record components is to apply a meaningful structure to constituent data items. Other international standards provide their own information or reference models and this work focuses on the EN ISO 13606 Reference Model, which represents the core component of the information governance knowledge model proposed in this thesis because it is the most generalisable example of such a model and can be adapted to represent information beyond that which is found in healthcare records. The ELEMENT is the only record component that holds actual data items. The data items supported by EN ISO 13606 include the data types in

the CEN Technical Standard CEN TS 14796 (European Committee for Standardisation, 2004), which are treated as attributes.

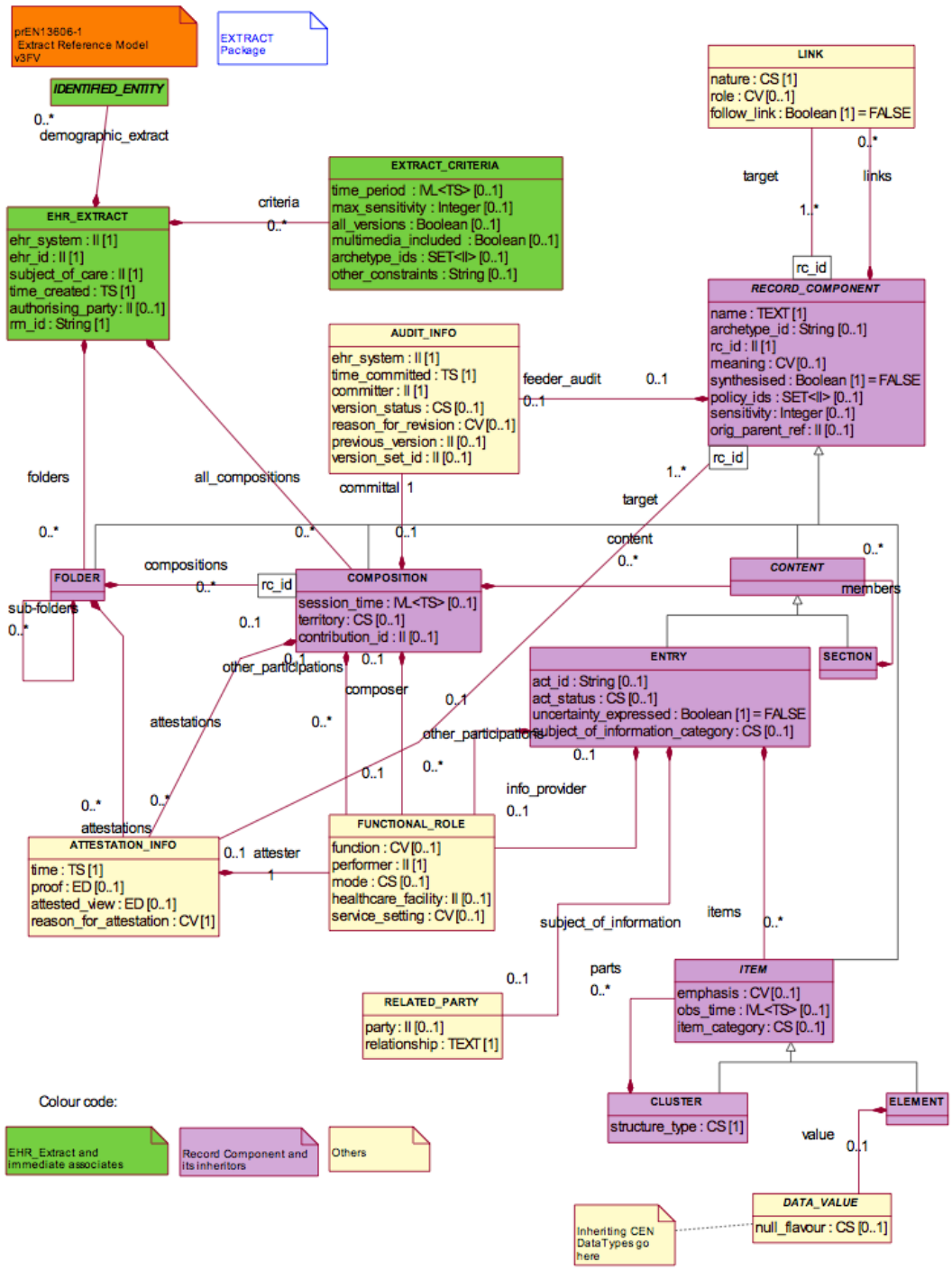


Figure 9: EN / ISO 13606 Reference Model (British Standards Institute, 2007a)

EHR HIERARCHY COMPONENT	DESCRIPTION	EXAMPLES
EHR_EXTRACT	The top-level container of part or all of the EHR of a single subject of care, for communication between an EHR Provider system and an EHR Recipient.	(Not applicable)
FOLDER	The high level organisation within an EHR, dividing it into compartments relating to care provided for a single condition, by a clinical team or institution, or over a fixed time period such as an episode of care.	Diabetes care, Schizophrenia, Cholecystectomy, Paediatrics, St Mungo's Hospital, GP Folder, Episodes 2000-2001, Italy.
COMPOSITION	The set of information committed to one EHR by one agent, as a result of a single clinical encounter or record documentation session.	Progress note, Laboratory test result form, Radiology report, Referral letter, Clinic visit, Clinic letter, Discharge summary, Functional health assessment, Diabetes review.
SECTION	EHR data within a COMPOSITION that belongs under one clinical heading, usually reflecting the flow of information gathering during a clinical encounter, or structured for the benefit of future human readership.	Reason for encounter, Past history, Family history, Allergy information, Subjective symptoms, Objective findings, Analysis, Plan, Treatment, Diet, Posture, Abdominal examination, Retinal examination.
ENTRY	The information recorded in an EHR as a result of one clinical action, one observation, one clinical interpretation, or an intention. This is also known as a clinical statement.	A symptom, an observation, one test result, a prescribed drug, an allergy reaction, a diagnosis, a differential diagnosis, a differential white cell count, blood pressure measurement.
CLUSTER	The means of organising nested multi-part data structures such as time series, and to represent the columns of a table.	Audiogram results, electro-encephalogram interpretation, weighted differential diagnoses.
ELEMENT	The leaf node of the EHR hierarchy, containing a single data value.	Systolic blood pressure, heart rate, drug name, symptom, body weight.

Table 1: EHR Hierarchy for a record extract, describing the Record Components

This allows for the specification of coding schemes, language support and null flavours across primitive data types, including Boolean, Byte, Character, Double, Integer, Real, String, Sets, Arrays and Lists, which are assumed by EN ISO 13606-1 to be available on any coding platform. These basic types are used to define the nature of the ELEMENT data value, combining the additional details and support structures for languages and code schemes to the ELEMENT data value within the wider structure of the record components.

To illustrate how the reference / information models work in the context of healthcare records, Figure 10 provides an example of how a diabetic patient presenting with head and back pain would have the clinical contact recorded according to the openEHR equivalent to the EN ISO 13606 Reference Model, using the Problem Based Patient Record approach proposed by Weed (Weed, 1968),

focussing on subjective and objective items, and assessment and a plan. In this example, the yellow and purple boxes relate to the information model structures including the COMPOSITION and SECTION described in Table 1, and the turquoise boxes represent the ENTRIES, which themselves contain the clinical data items according to the constrain model specified by an Archetype.

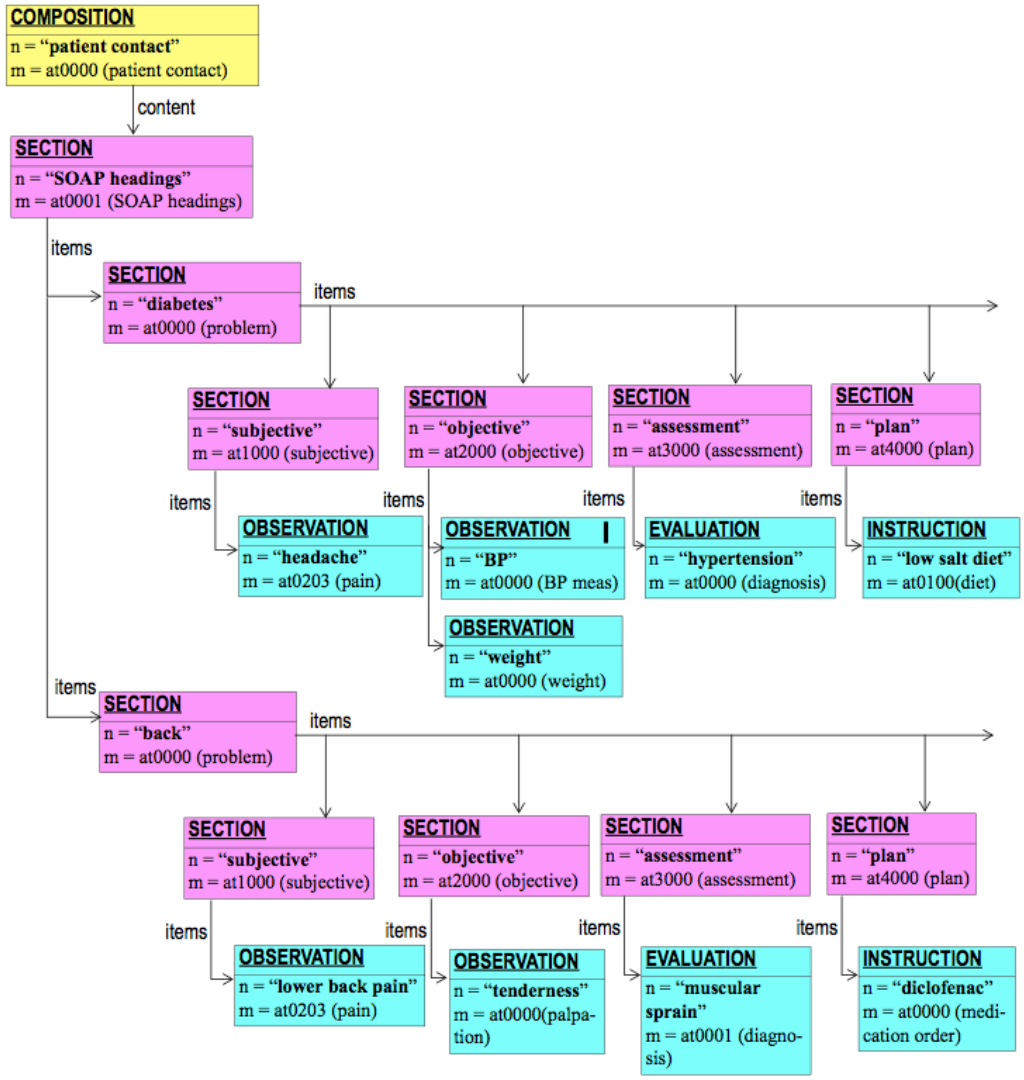


Figure 10: Example of openEHR Content package in use

Using these Reference / Information models, a structure can be applied to record components within the EHR, which are further categorised and arranged according to the constraints specified by the Archetype and Template approach.

4.4.3. Constraint Modelling: Clinical Archetypes

The structure of an EHR is an important aspect of semantic interoperability and consistent understanding. The concept of the *Template* in HL7 (Health Level 7 (HL7), 2014b) or an *Archetype* in openEHR (Leslie, 2012) and EN / ISO 13606 is recognised as a "blueprint" or constraint model for clinical concepts (Garde et al., 2009) core to semantic interoperability (Garde et al., 2007b, Tapuria et al., 2013, Maldonado et al., 2012). Although Bointner and Duftschmid identify incompatibility between Templates and Archetypes (Bointner and Duftschmid, 2009), the concept of a structure or constraint model remains consistent. The Archetype represents the knowledge artefact that can be authored, reviewed and edited by members of the clinical community as they seek a consensus on how to represent commonly used clinical concepts, for example a Blood Pressure. The Archetype handles the representation of clinical concepts by using constituent information items represented in the Reference or Information models to define the precise structure and permissible values of information captured about an individual during their healthcare interactions. This constraint modelling is designed to specify which information components belong within a particular clinical concept, and how they interrelate with each other. The Archetype can be used to bind not only information and reference models, but also clinical terminologies (Berges et al., 2014, Meizoso Garcia et al., 2012, Qamar and Rector, 2007), which help to provide domain experts a means by which they reach a consensus on how to model a clinical concept and give a clear indication of what is being represented.

Using the openEHR approach by way of example for a blood pressure, this would be represented as a subtype of ENTRY (for example, the BP OBSERVATION as provided in Figure 10) in a clinical record. A blood pressure usually involves several data items that hold the information, including a systolic measurement, a diastolic measurement and a pulse, where some examples would include the position of the patient (i.e. sitting or standing). The Archetype specifies the constraints on the representation of a Blood Pressure: by referring to the example in Figure 10, it would specify that it would appear in a patient contact

COMPOSITION, and would be subcategorised under three SECTION headings according to the SOAP headings and within the diabetes "problem" as an objective measurement. The Archetype would further specify the name of the observation (i.e. "BP") and would further allow for several items of information (comparable to the ELEMENT in EN ISO 13606), including the DATA parts that would include the systolic reading, the diastolic reading and the pulse, and a STATE parts, which would describe whether the patient was sitting or standing. The Archetype would specify further constraints on the DATA parts, for example, that the systolic and diastolic readings should be within a range of zero and three hundred, and that the systolic should be higher than the diastolic. These constraints would also apply to the STATE part, allowing only the values of "sitting" or "standing" to be used. Figure 11 illustrates the openEHR class diagram for the Archetype model, and Figure 12 the EN ISO version. specified using the *assertion* package, whilst the *constraint* package specifies the configuration of the Reference Model classes (i.e. COMPOSITIONS, SECTIONS, ENTRIES and ELEMENT leaf nodes) through the C_COMPLEX_OBJECT package, illustrated in Figure 11.

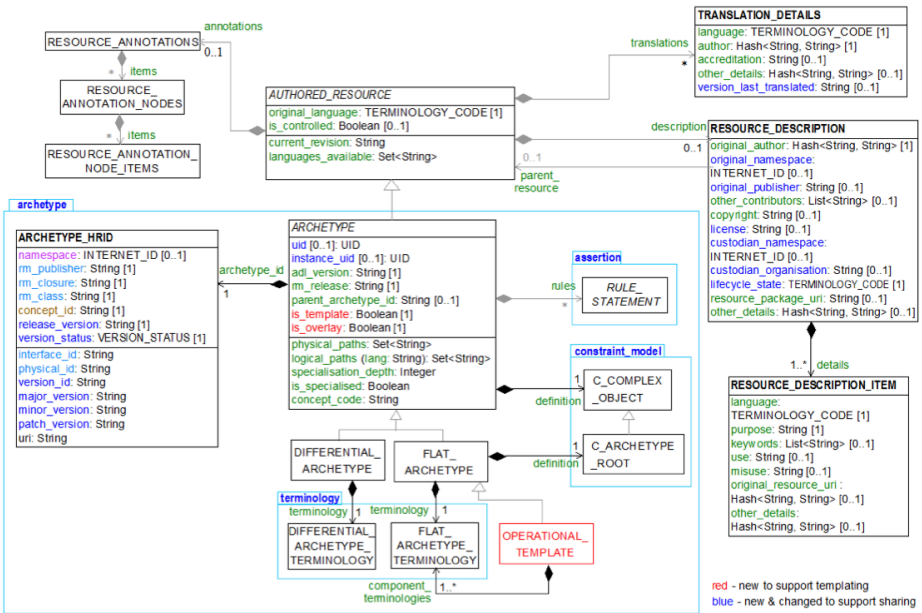


Figure 11: Archetype Model as defined by openEHR (openEHR, 2014a)

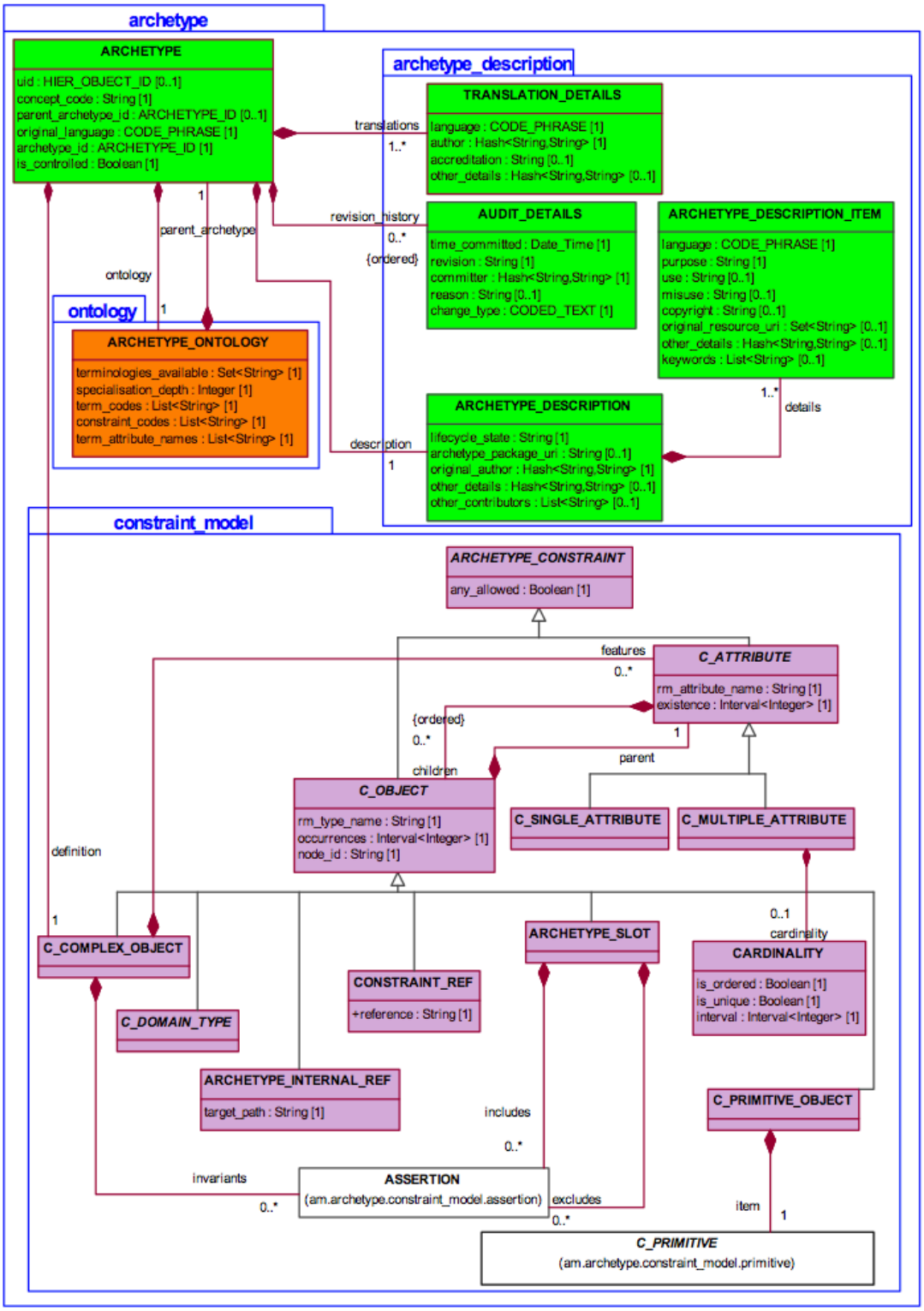


Figure 12: EN / ISO 13606 Part 2 Archetype and Constraint Model (British Standards Institute, 2007b)

The constraints internal to the Archetype (for example on data values) are The EN ISO 13606 Archetype model shown in Figure 12 places the C_COMPLEX_OBJECT

class in the class structure within its Constraint package and includes an *assertion* package, combining the constraint specifications for both the EN ISO Information Model classes and the constraints within the Archetypes on the ELEMENT item data values. Both figures show the details that are captured to describe the Archetype, the relationship between Information or Reference models (described in section 4.4.2) and terminologies, coding schemes and terminologies. The Archetype integrates the reference / information models and terminology models, with support for ontology representation. Coding schemes, terminologies and ontologies are described in section 4.4.4.

A number of editing tools exist to enable the authoring or reviewing of archetypes that are compliant with the EN / ISO 13606 standard, or openEHR. Examples include the Object Dictionary editor as implemented by Austin (Austin, 2004), and the ADL Workbench (openEHR, 2014d), which uses a formal language for the specification of Archetypes called the Archetype Definition Language (ADL) (openEHR, 2014c) to allow the specification of Archetypes. These Archetypes can then be reviewed, commented upon and edited by collaborating members of the clinical community: openEHR has developed a library of clinical archetypes, which it proposes can be reused across clinical practice (openEHR, 2014f). This is aimed at limiting the need to author Archetypes for clinical concepts that had already been authored by other teams in the spirit of open collaboration.

ADL underpins an online and review tool, the Clinical Knowledge Manager (CKM), which facilitates this collaborative working (openEHR, 2014e). This approach has been used to facilitate harmonisation between clinical users and the development of EHR Systems, where work has been underway to integrate support for multiple Reference and Information Models (Maldonado et al., 2009). Austin et al. have provided a similar approach to harmonisation according to an XML schema (Austin et al., 2013), where Rinner et al. have attempted a validation of an EHR schema using XML (Rinner et al., 2010) and Sanchez-de-Madariaga et al. have proposed new markup languages to represent ISO 13606 Extracts (Sanchez-de-Madariaga et al., 2013).

4.4.4. Clinical Coding, Terminologies and Ontologies

All the major standards make allowance for binding specific terminology and coding libraries to provide a label for clinical concepts according to a commonly agreed coding scheme, as has been described in section 4.4.2 and can be seen in Figure 12. These schemes code clinical concepts, including diagnoses, procedures and clinical events. The International Classification for Diseases (ICD) is one example, which handles diagnostic codes in epidemiology, clinical practice and healthcare management (World Health Organisation (WHO), 2014). The International Health Terminology Standards Development Organisation (IHTSDO) has also developed a terminology set for clinical terms specifically for EHRs called the Systematized NOMenclature of MEDicine Clinical Terms (SNOMED CT) (International Health Terminology Standards Development Organisation (IHTSDO), 2014) found in hospital and GP EHR systems. The NHS maintains a data dictionary thorough the HSCIC, which refers to additional coding schemes including the former Office of Population Censuses and Surveys (OPCS) operating procedure codes of interventions and procedures (Office of Population Censuses and Surveys (OPCS), 2014). Other examples of coding schemes and terminologies, and the Unified Medical Language System (UMLS) has been developed to integrate these terminology, classification and coding schemes to "…promote creation of more effective and interoperable biomedical information systems and services, including electronic health records…" (US National Library of Medicine, 2014).

These terminology and coding schemes are used to explicitly label components of an EHR so that people can review records and unambiguously identify the clinical concept being described during care provision and research uses of records. Integration of clinical coding schemes with the Archetype approach is essential to semantic interoperability. Sundvall et al. describe one such approach, where they present an archetype editor equipped with support for manual or semi-automatic creation of bindings between archetypes and terminology systems based upon a series of suggested terms for automatic mappings, a process assisted by visualisation tools to help with the mapping and term exploration (Sundvall et al., 2008).

Where terminology provides a series of terms and concepts that have been coded according to clinical coding schemes, ontologies provide a means to represent computable, conceptual models and semantic relationships, providing a "…clear abstraction of knowledge from the underlying operational applications that will make use of them…" (Corrigan et al., 2013). W3C has developed a web publishable ontology called the Web Ontology Language (OWL) (Worldwide Web Consortium, 2012), which has been used to develop a richer knowledge base by mapping the EHR Archetype and Reference models to ontologies and coding schemes in the healthcare record area (Lezcano et al., 2011, Martinez-Costa et al., 2009, Santos et al., 2010). OWL has been used to integrate knowledge management through ontologies in daily clinical practice (Griffon et al., 2013) and to help develop decision support tools (Minarro-Gimenez et al., 2014, Bau et al., 2014). It has also been useful for integrating and enriching data sets for research and guiding the development of clinical queries (Chen et al., 2014, Liang et al., 2014).

4.5. Technologies to Implement EHR Servers

The Archetype, Reference and Ontology components represent the elements of knowledge management for EHRs and must be integrated effectively in order for users to be able to review and understand the information that they hold within a logical EHR. The various components are developed using a series of technologies, including web based technologies for presenting knowledge artefact editing tools, clinical forms and records to users, high level languages for developing business logic that helps to process the information and apply constraints to how the information is structured, and some form of database to store the clinical, record structure and modelling information, including the labelling and coding schemes discussed in the previous section. This section describes the technologies used to implement and deploy knowledge managed information systems used to provide healthcare services and support clinical research. Figure 13 offers a view from openEHR regarding the development components and deployment architecture for developers and EHR system vendors, built upon the health informatics infrastructure (openEHR, 2014g).

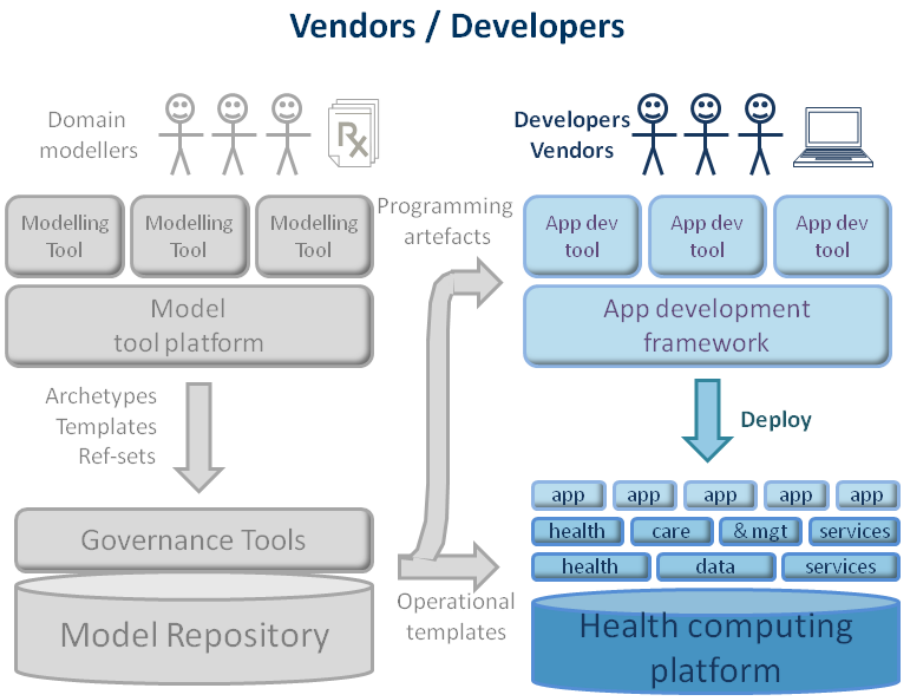


Figure 13: Overview of Knowledge Management and EHR system components (openEHR, 2014g)

EHR servers generally fit the paradigm of the *n-tiered architecture*, which organises the different components of a system into several tiers. Typically there are three tiers, a presentation, business logic and persistence tier. The presentation tier encapsulates the frameworks and implementations that are needed to provide a user interface, whilst the business logic tier provides the implementation of the business rules for handling data, applying modelled constraints and allowing it to be stored and retrieved from the persistence tier, where raw data is held, usually by a database management system.

EHR systems have been developed and deployed using a variety of technologies. Web applications have become one of the more prominently used examples for managing EHRs because of their tiered approach to deploying information processing components, which broadly fit the n-tiered paradigm. Kobayashi et al. describe the implementation of an EHR web application system and refer to implementations using C# and Java, which allows presentation tier elements to be generated based upon the Archetype models that have been prepared (Kobayashi et al., 2013). This approach is consistent with methods observed during the case studies described in Chapter 5, where a combination of

web page, business logic and database management system technologies were observed and used by the author for development.

4.6. Implementing Policy-Based Controls

The structure of EHRs provides a basis for capturing rich, detailed information about healthcare interventions. This information is subject to the information governance requirements described in Chapter 3 and work has been carried out to develop methods and approaches to protect them. This section describes the mechanism that are in place to protect the records, which include Privilege Management and Access Control (PMAC), and a series of examples of computable policy-based control specifications and formalisms that have been considered for EHR systems.

4.6.1. Controlling Access and Privileges for EHR Access

Regardless of the risks that individuals will be re-identified from anonymous data or whether and how consent from participants must be sought, access to research data sets must still be controlled. Access controls defined by policies and based upon the role that a user holds within an organisation (Sandhu et al., 1996) is a preferred solution to meet privacy and confidentiality protection requirements for EHR systems (Blobel, 2004, Ferreira et al., 2007, Lin and Brown, 2000, Carrion Senor et al., 2012, Motta and Furuie, 2003, Alhaqbani and Fidge, 2008). The literature recognises that no one basis is sufficient to model access policies and privacy requirements (Al-Fedaghi, 2007) and that context is an important factor for controlling access to healthcare records, referring to other bases for access control, including privacy oriented access models for EHRs (Gajanayake et al., 2012), where discretionary, purpose based and mandatory access are explored. Sandhu et al. emphasise that access control must be regarded as complimentary to other forms of security policy control, including audit (Sandhu and Samarati, 1994) where an audit basis for controlling access has been developed (Dekker and Etalle, 2007).

Work performed on the area was realised in an ISO Standard, EN/ISO 13606 part 4 (British Standards Institute, 2009), which proposes privilege management

as a means to augment roles based access control for the electronic healthcare records. The role is assigned a privilege level that is mapped to a sensitivity level, which in turn has been assigned to certain classes of information. An example provided in EN ISO 13606-4 is that of an HIV test result for an individual, which would be assigned a high sensitivity level because of the ramifications for the individual if it were liberally disclosed; only an attending clinician with a legitimate relationship of care, or a researcher who had been granted approval by an ethics committee to view that information, would be assigned the appropriate role, which in turn determines what sensitivity level they can access. This example is described in detail in section 7.3.5. Even if the record of the individual were rendered anonymous, the sensitivity of the de-identified information would still need to be considered by a research project's steering committee along with the agreement of the releasing organisation and any additional ethics committee stipulations.

An example of a measure to allow the restriction of access to sensitive information was the Sealed Envelope, illustrated by Becker when he argues the case for using formal expression of policy in the context of clinical care, citing examples where patients have deemed that certain data about them is sensitive and should be concealed from general access: the Sealed Envelope aimed to permit access to users that have held a legitimate relationship of care with that patient and need to see the data to provide that care (Becker, 2007).

4.6.2. Using Computable Policy Specification

The modelling of policy for access control and privilege management is clear in the literature. This is illustrated by Blobel (Blobel, 2007), Lin et al. (Lin and Brown, 2000) and Ferreira et al. (Ferreira et al., 2007). Additionally, access policy has been expressed using formal representations of required policy stipulations, including the PONDER policy specification framework ((Sloman and Lupu, 2002), CASSANDRA and SP4 (Becker, 2005, Becker et al., 2010). The OASIS Foundation has develop a series of open standards (Organization for the Advancement of Structured Information Standards, 2011c, Standards, 2011) that specify rules for access control (Organization for the Advancement of Structured Information

Standards, 2011b, Organization for the Advancement of Structured Information Standards, 2011a). IBM also offered EPAL for privacy controls (IBM, 2003).

There are identified issues with these approaches. Protection challenges that are posed to the processing of clinical data have been illustrated by the literature review performed by de Lusignan et al. (de Lusignan et al., 2007), which emphasises two consistent themes across the literature: comprehensive security management facilities are lacking and many of the solutions are untested outside of a laboratory environment. A third theme is that of the risks of error in and complexities of policy interpretation, as highlighted by Becker (Becker, 2007), who illustrates the issues of refining a policy that is human readable to one that is computable at a software run time level within several different software components. He discusses problems that ambiguity and incompleteness pose in a policy expressed in natural language and also proposes that RBAC would not cope with the complexity that NPfIT healthcare systems required based on information governance policies.

4.7. Summary of Information Strategy and Support for Clinical Research

This chapter has described the key strategic drivers for the development of information management strategies to support healthcare provision in the UK and beyond. It has discussed the development of information technology resources to support care and the ability to share healthcare record information for other purposes, including clinical research. It has described the support that has been provided by UK and European research councils to develop databases of information that have been derived from electronic healthcare record systems, which have captured details from individuals as they have presented for care services. This chapter has also explored the core standards, which have been ratified internationally and implementations that have underpinned electronic healthcare record development, and the main aspects of knowledge management that enable the effective sharing of information for care and research.

It is clear that there is significant governmental and research council support for sharing healthcare records for research, a significant portion of which has been

provided to the development of computing resources and the storage of electronic healthcare information. This has continued despite the anxieties described in section 3.3. It is also clear that the risks to participants, including re-identification, are considered balanced by the advantages of the research projects, whether they are smaller scale cohort studies, focussed on individual benefit or aimed at population scale, epidemiological studies. The support for linking records either for care or through research data sets is significant, and these elements have been integrated into the strategy for information management in the UK, Europe and the US, as well as beyond.

There is a compelling case to ease the management of not only the information assets, but also their protection. Whilst there is clearly a wealth of policy-based control mechanisms available that lend themselves to formal specification or scripting, particularly for access control and privilege management, but these do not integrate for common knowledge representation or human readership and must be manually adapted to some of the governance requirements and used within the framework of other control solutions, including overall policies and information governance framework components.

The specification of electronic healthcare record structures to accurately model the clinical concepts that they represent is rigorous and is supported by a comprehensive set of engineering capabilities, but this process is complex. The core record models are specified according to a dual modelling approach, that allows for the specification of a record structure based on a series of record components, and a constraint model for how a record should be structured to represent the clinical concept about which information is being stored. By integrating clinical terminologies, this approach can provide a consistent view across the care and secondary use teams, both of which are increasing in number along with the uses of the information. This approach supports knowledge management framework development, which relies on semantic interoperability and is itself hard to achieve.

This dual modelling approach has allowed the author to consider and develop a knowledge management solution for managing information governance requirements by developing a generic model that represents information

governance concepts and deploying web application tools to provide users with a consistent view of what is expected of them when working with electronic healthcare record information. The next chapter describes the research project and clinical system deployment case studies that helped to develop the knowledge management framework for information governance requirements. These gave the author an opportunity to understand better the use of the knowledge management approach in the clinical domain, and to adapt a simplified approach that could model information governance concepts and deploy them for guiding people on how to behave with sensitive healthcare information when used for research.

Chapter 5. Research Project Case Studies

The sharing of personal health information has become increasingly commonplace for the provision of health care services internationally, as well as a valued means to support medical research, disease surveillance and service delivery improvement. Information sharing has become easier and more accurate because of a number of factors, including the rapid evolution of cheaper, more widely available computer hardware, more sophisticated knowledge management software and the provision of faster networking that promotes collaboration and resource sharing. The sharing of this information has posed challenges to meeting the legal, ethical and practical requirements for achieving effective information governance. These challenges include the need to interpret a wide range of legislative, standards based and good practice guidelines so that EHRs can be handled and used appropriately. This relies on the development of polices and risk mitigation strategies that also need interpretation and understanding about what is expected in terms of behaviour and appropriate practice when conducting research and handling the EHR data.

The author has observed these challenges directly whilst working on a series of research and information platform development projects, which have provided an opportunity to perform case studies around the use of sensitive data for clinical research. Since the data has been released to these projects after being de-identified to varying degrees, there still remained in each case the risk of re-identification and the information sometimes had to be regarded as confidential and sensitive. The case studies involved collecting information on working practice, information flows, data use and engagement with key stakeholders in these research projects. This helped to understand the issues that were observed in practice and helped to guide the requirements collection to develop a knowledge management framework. The goal of this framework was to clarify the expectations of good practice and information governance requirements for users, help encourage a consistent interpretation and understanding of expected behaviour and encourage good working practice by helping users apply their expertise when working with sensitive information. This chapter describes each of

the research project exemplars that the author has worked on, the experience and results of which have helped to shape the development of the knowledge management framework. Each section describes a case study, the methods used to gather information and develop the framework.

5.1. Clinical eScience Framework and the eScience Initiative

The experience of working on the Clinical eScience Framework (CLEF) Project and engaging with the eScience Initiative presented an opportunity for the author to observe the processes, provisions and expectations for information governance and security management at a practical level when using health care data for research purposes. Sensitive data use in medical research had become topical within the research community in 2003 and 2004, mainly due to the difficult experience that both research projects and NHS research ethics committees encountered when computing researchers were about to start using sensitive data in their work, and ethics committees were starting to review projects that were heavily based on computational disciplines: the parties responsible for pursuing the governance process were presented with challenges for which there had been little experience or understanding of the issues at hand. These issues had been the topic of discussion at workshops and symposia discussed in the first chapter (Oxford Internet Institute, 2005, Royal College of Practitioners, 2004), which the author attended. Figure 14 gives an overview of the information flows and protection points for the project (Kalra et al., 2005).

5.1.1. Information Governance Focus for CLEF

CLEF focused on policy development, de-identification of records and the development of a policy framework in accordance with the ethics committee requirements. It acknowledged the experience of other eScience projects, and made a concerted effort to ensure that all required governance details were available and applied. The result was that security management became a top priority for the project, and a work package was dedicated to this area. The day-to-day management of the security requirements were based upon ethics

committee stipulations that specified the need to maintain the anonymity of patients and that all data be destroyed at the end of the project.

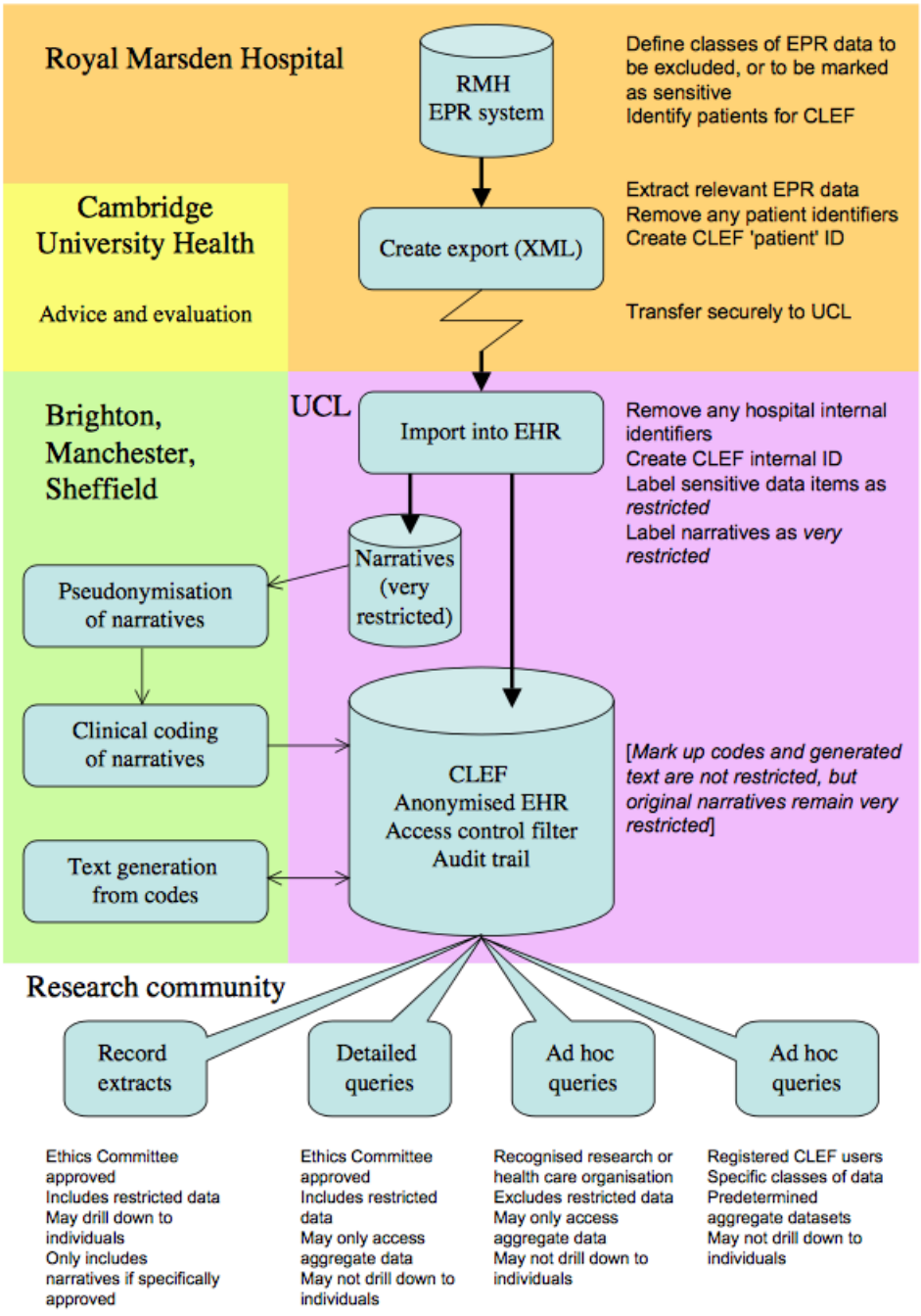


Figure 14: Overview of information flow and protection points for CLEF (Kalra et al., 2005)

This prompted a full review of available governance and security standards, good practice guidelines and legislative artefacts for the management of these requirements, which have formed part of the literature review discussed in sections 3.1 and 3.2. As the requirements for the security management were

established, the project recognised the complexity involved and the caution with which it would have to proceed. The author helped to develop policies to encapsulate these requirements and expectations, which are provided in Appendix 1. The author took this opportunity to make observations of the research staff and their working practice, which showed that detailed knowledge was required to comprehensively assert the requirements and provide assurance, but that there was no provision to manage that knowledge beyond specifying human readable policy documents, or use it in a timely fashion to assert the requisite security controls in a computed environment.

In addition to the observational work performed within CLEF, engagement with the eScience Initiative provided further opportunity to explore the research area. The Initiative supported a branch of eSocial Science that reviewed the ethical and governance issues that were being raised. The author's attendance at the workshops and symposia provided an opportunity to analyse further motivating factors behind research governance and wider concerns about protection of sensitive information in the health care context from both a practical and ethical point of view, which are described in section 3.3. There was a clear need not only to consider the use of policy-based controls, but also to assist with the interpretation of the governance process, reflection of that interpretation in written policy and assertion of the resulting policy stipulations. This understanding was developed through stakeholder interviews, described in the next section.

5.1.2. CLEF Stakeholder Engagement, Meetings and Interviews

This set of observations and involvement with the eScience community formed the initial motivations behind re-applying the EHR knowledge management approach to develop a solution to the observed problems of information security management in the context of sensitive data reuse in medical research. A series of informal project meetings and interviews yielded further insights into the issues of managing information governance for a research database. This helped provide a fuller understanding of how individuals and professionals perceived the protection expectations and requirements, and to explore areas that might require assurance,

particularly when using potentially identifying data for purposes such as medical research. Figure 15 lists the stakeholders that CLEF researchers identified for research uses of healthcare records and the wider eScience agenda, where the majority of the meetings and interviews were held with the internal stakeholders.

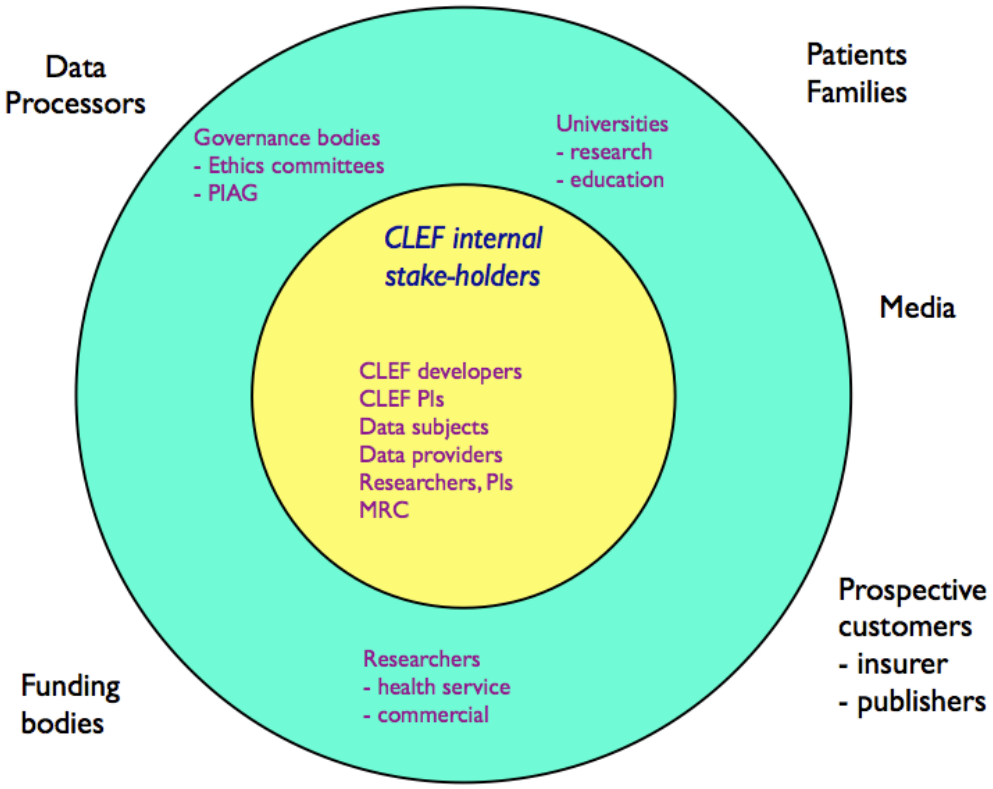


Figure 15: Internal and External Stakeholder Groups

The meetings and interviews were minimally structured to allow for a broad range of qualitative responses as the research problem area was explored. Discussions with the researchers and Principal Investigators helped the author to understand what kinds of information would be needed to answer clinical research queries. The researchers included computer science and natural language experts, data management and ontology researchers, and clinical researchers and consultant oncologists.

Engagement with the developers in CLEF helped to gather an understanding about the software that they were building to process potentially identifying, sensitive data. This software included the integration of an existing information security management framework that combined role based access control software developed for eScience projects called PrivilEge and Role Management

Infrastructure Standards (PERMIS) (Chadwick and Otenko, 2003) and the Flexible Authentication Middleware Extension (FAME) infrastructure, which focused on authentication (Zhang et al., 2004), and this resulted in the FAME-PERMIS infrastructure (Zhang et al., 2007). The collaboration with in CLEF provided an opportunity to ask developers to define what roles existed in the processing of the data, what privileges were required and who in the team of researchers would be processing what component of the software. The meetings sought also to define the minimum data access requirements – specifically what was the minimum set of data that needed to be released (or that a user would need to access) in order to perform the specified processing tasks. This also allowed the author to gain a better understanding of the scope of the FAME-PERMIS solutions and how best to adapt a representative medical research data repository to be configured with this software. This needed the interpretation of information governance requirements so that a fuller understanding of the extent to which the authentication and authorisation components would meet the overall expectations could be achieved. The roles and access privileges that were defined for CLEF through this work are available in Appendix 2.

Further interviews were conducted with data providers, which helped to develop understanding about the complexities and risks of transferring information outside of the care setting and into the research environment, particularly regarding the risks of re-identification. The combined set of interviews and requirements gathering helped to understand the flow of information, de-identification and access and authentication policy enactment presented in Figure 14. The interviews also focused on an understanding of defining the risks of re-identifying patients from de-identified records using a method of Statistical Disclosure Control (SDC) (Smith and Elliot, 2008), and audit requirements for a research data repository like the one constructed for CLEF, which is illustrated in Figure 16, which was developed by Dr Brock Craft and Mr Peter Singleton. The need for assurance mechanisms and importance of protection was emphasised in interviews with privacy experts who formed part of the CLEF team.

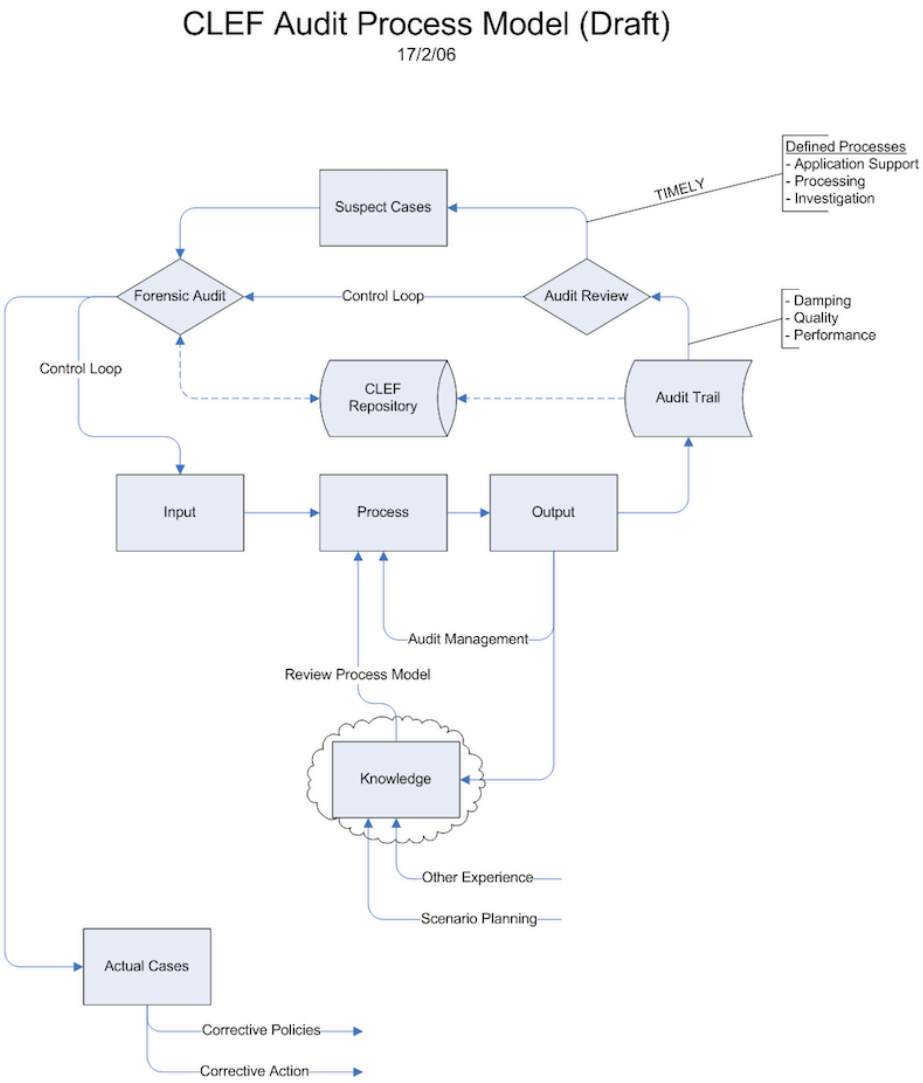


Figure 16: Audit Process Model

CLEF and the eScience Initiative provided a foundation for the research area. They illustrated the benefits of reusing clinical information for research purposes and the potential for providing answers to clinical research queries and translational research that benefitted care directly. They also clearly illustrated the complexities of achieving information governance and the level to which the scope of information governance was not fully understood by the clinical, research and legislative community. It was also clear that the stakeholders involved were not fully conversant about how their rights and interests would be upheld when information was reused for healthcare, or the basis upon which ethical review could occur, which was one reason why deceased patient records were used for

CLEF. The eScience initiative helped to illustrate the extent to which the approvals process for ethical review and governance definitions was too slow, prompting the research ethics committees to commit to a forty day turnaround for reviewing applications as opposed to the eighteen months that it would sometimes take them. CLEF and eScience both showed that there was a need for engagement, education and communication regarding the research strategy. Working on the project showed that achieving effective information governance needed a tighter co-ordination between the legal and ethical foundations for protecting information and the good information security practice that would put them into effect. It also demonstrated a clearer way to engage with the people who were governed by the policies and expected working practice guidelines that were developed.

5.2. Databases for HIV: Integration, Collaboration and Engagement

The Databases for HIV: Integration, Collaboration and Engagement (DHICE) Initiative was a Wellcome Trust / MRC funded project that researched a number of issues with the collection of HIV data for both medical research and surveillance activities (University College London, 2010b). The Initiative was formed of eight cohort and population studies that collected EHR data from Genito-Urinary Medicine (GUM) and HIV clinics, focusing on aspects relating to treatment outcomes using various anti-retroviral therapies, including antibody (CD4) levels and viral load. In contrast to CLEF, this case study featured a series of cohort and population studies that were already collecting data separately but were working collaboratively to answer their common clinical research queries and enrich the available information using a federation of databases that held records about individuals at different ages and treatment centres around the UK. The goals of the Initiative were to improve the data flows from the specialist treatment centres to the studies, expand the existing collaborations between the studies under a common information governance framework and to engage the HIV treatment advocate community to help them understand the research that was being performed and how information governance was used to protect participant records.

5.2.1. The DHICE Studies and Collaborative Research Focus

The partner studies in the Initiative were existing, separately funded projects that focussed on research into HIV and treatment outcomes across different cohorts. These included the surveillance of HIV prevalence in the UK, development of treatment strategy and management policy, commissioning of services and maintaining a registry of new diagnoses. One of the studies involved was the UK Collaborative HIV Cohort (UK CHIC) (UK Collaborative HIV Cohort Study, 2014), which researched the effects of treatments on approximately forty per cent of the HIV population, for example the effects of late diagnosis on life expectancy (May et al., 2011). Another study was the National Survey of HIV in Pregnancy and Childhood (NSHPC), which collected data about pregnant women diagnosed HIV positive and children who were born HIV positive, and other HIV positive children (University College London, 2013). An example of research performed by NSHPC includes the responses of HIV positive pregnant women to antiretroviral therapies (Huntington et al., 2014) amongst others. To support common research goals NSHPC collaborated with UK CHIC on a number of occasions, formulating a method of linking records in the two studies (Huntington et al., 2012).

UK CHIC and NSHPC also collaborated with the Collaborative HIV Paediatric Study (CHIPS) (Foster et al., 2009), which is a collaboration between the Paediatric European Network for the Treatment of AIDS (PENTA), NSHPC and the MRC Clinical Trials Unit studying HIV in children and adolescents (Collaborative HIV Paediatric Study, 2014). In addition to these, a register of HIV Seroconverters (Medical Research Council, 2014c), which maintains a register of HIV positive individuals who went from an HIV negative diagnosis to HIV positive within six months of the initial HIV test also collaborated with DHICE. The UK Register also worked with UK CHIC (UK Collaborative Group on HIV Drug Resistance et al., 2007) and the HIV Drug Resistance Database (UK HIV Drug Resistance Database, 2014), which performed research on virus strain gene sequences and resistance from blood samples collected at laboratories around the UK. The Health Protection Agency (now Public Health England) was also part of the Initiative: it ran three population scale studies for new diagnoses, a surveillance of CD4 results

and the Survey of Prevalent HIV Diagnosed (SOPHID) projects, all of which were used to inform research, develop policy and determine payments for HIV centres around the country based on how many HIV positive individuals were being treated. The three studies were combined into the current HIV and AIDS Reporting System (HARS) (Public Health England, 2014). This system also received data from NSHPC and CHIPS.

The DHICE Initiative included collaboration with HIV treatment advocate networks, including i-base (i-base, 2014) and the UK Community Advisory Board (UK CAB) (UK Community Advisory Board, 2014). These networks helped to support the HIV community and provide information, guidance and advice for those living with and treating the condition. Members of these networks also sit on the UK CHIC steering committee and their work with DHICE was to engage with the research community so that they could better understand the research that was being performed, the studies that were being run and how records were being held and protected to support the cohort and population studies. This afforded the opportunity to meet, discuss and interview a set of research and community advisory stakeholders to achieve the Initiative goals and further develop requirements and tooling for an information governance knowledge model.

Working towards achieving the Initiative goals formed the basis of this case study. The work focused on improving information flows from clinical sites to the study databases and both developing existing and setting up new record linkages between the cohort and surveillance studies to enrich the records across the studies and provide more information upon which to continue the individual and collaborative research work. This required a review of the information governance framework that existed for each of the studies, the requirements for setting up new linkages and the development of a common information governance framework that would work across the studies as they continued to collaborate. Another goal of the Initiative was to engage with the treatment advocate groups so that they could better understand the research work that the partner studies were performing, find out more about how information was processed and thereby allow them to articulate the community concerns about privacy and confidentiality when information was being used for research.

Part of the review and engagement with the treatment advocate networks was to develop a visualisation tool that succinctly described the information flows and what they were being used for to help the public and HIV community understand what research was being performed. The UK CHIC, NSHPC and CHIPS information flows are illustrated using this visualisation tool in Figure 17, Figure 18 and Figure 19 respectively.

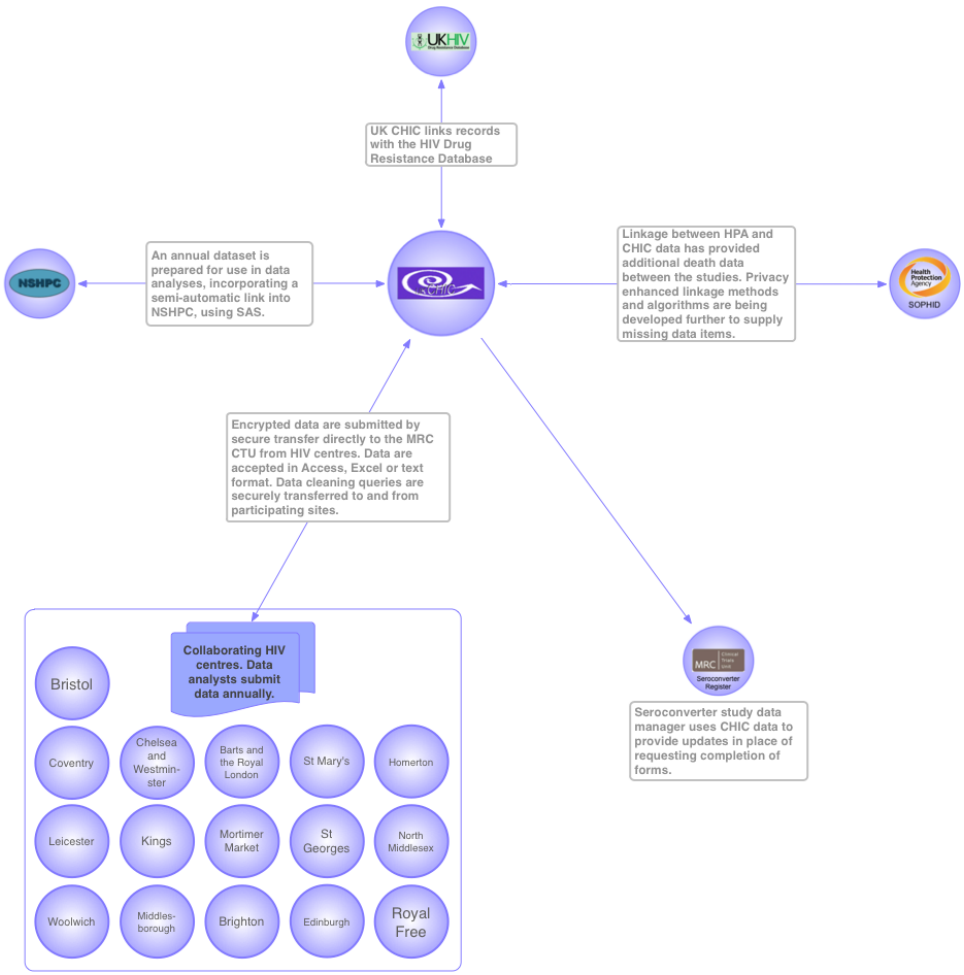


Figure 17: UK CHIC Information Flows and Linkages

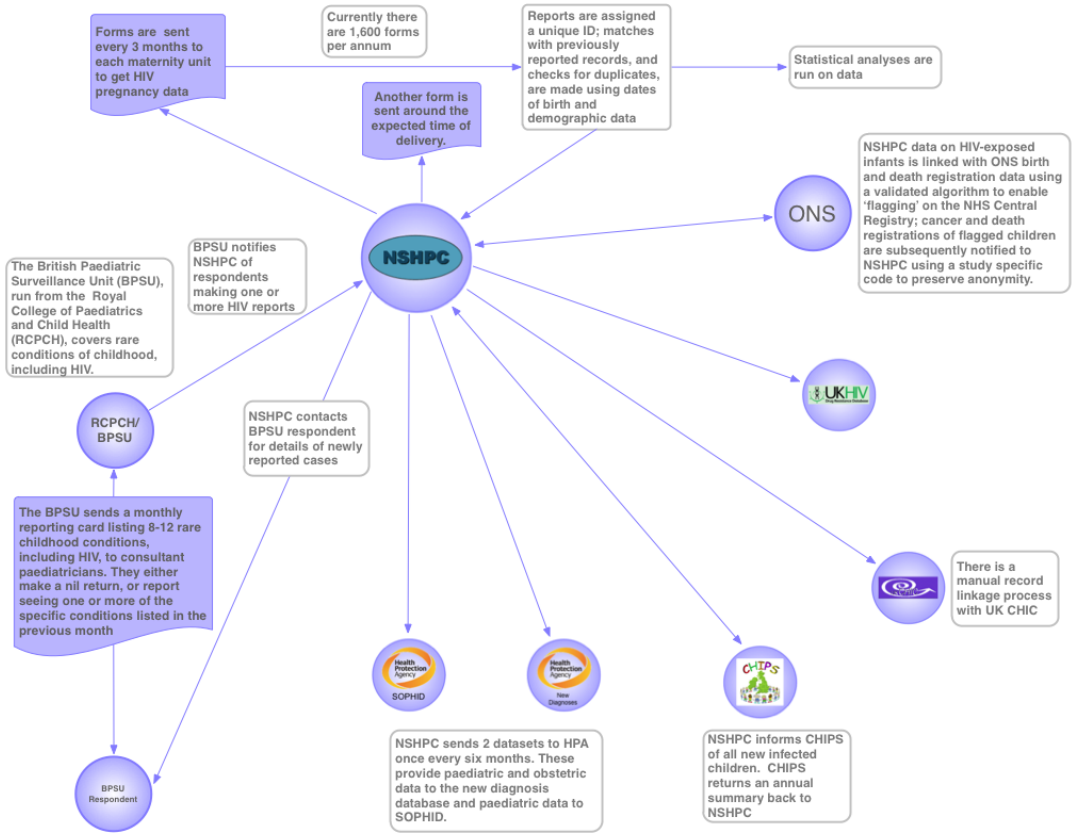


Figure 18: NSHPC Information Flows and Linkages

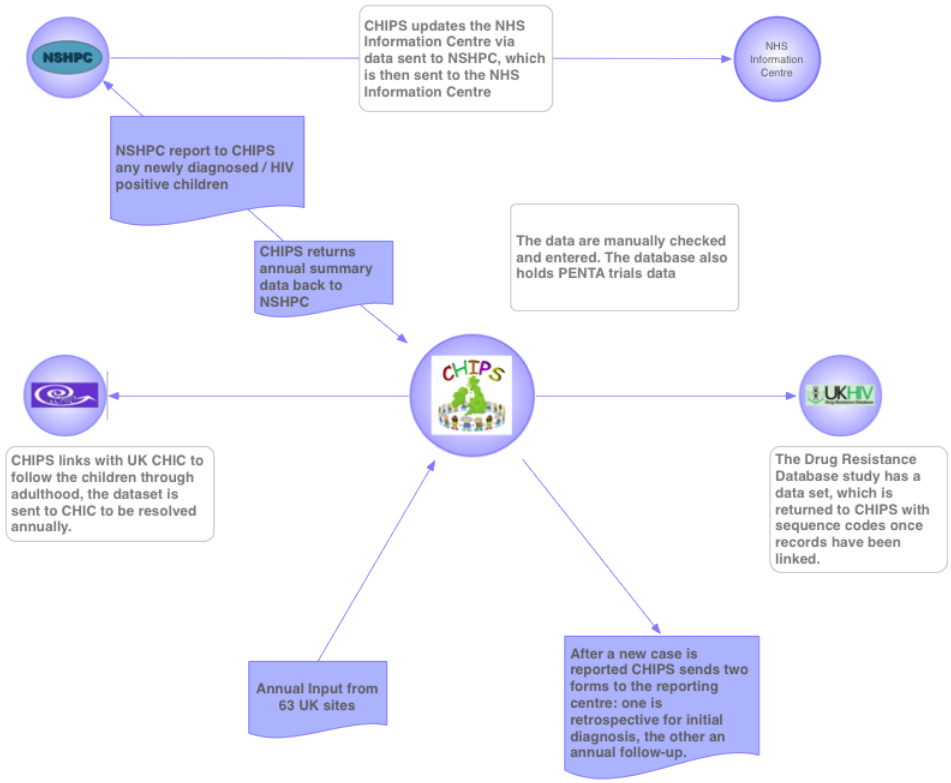


Figure 19: CHIPS Study Information Flows and Linkages

Understanding these data flows helped to get a better awareness of the information governance stipulations that were in place and the existing linkages that had been established between the studies. It also helped to define the linkage strategy that was needed to build a common data repository that would allow each of the studies to share and update their databases with agreed data items, as illustrated in Figure 20. This in turn helped the development of a common policy framework for each of the studies to use so that they were working to a consistent information governance template, which is provided in Appendix 3. This was a necessary step to try to incorporate the existing wealth of information governance policy and develop a means to refer back to existing policy documentation so that each project had a consistent template for specifying their information governance requirements and expectations.

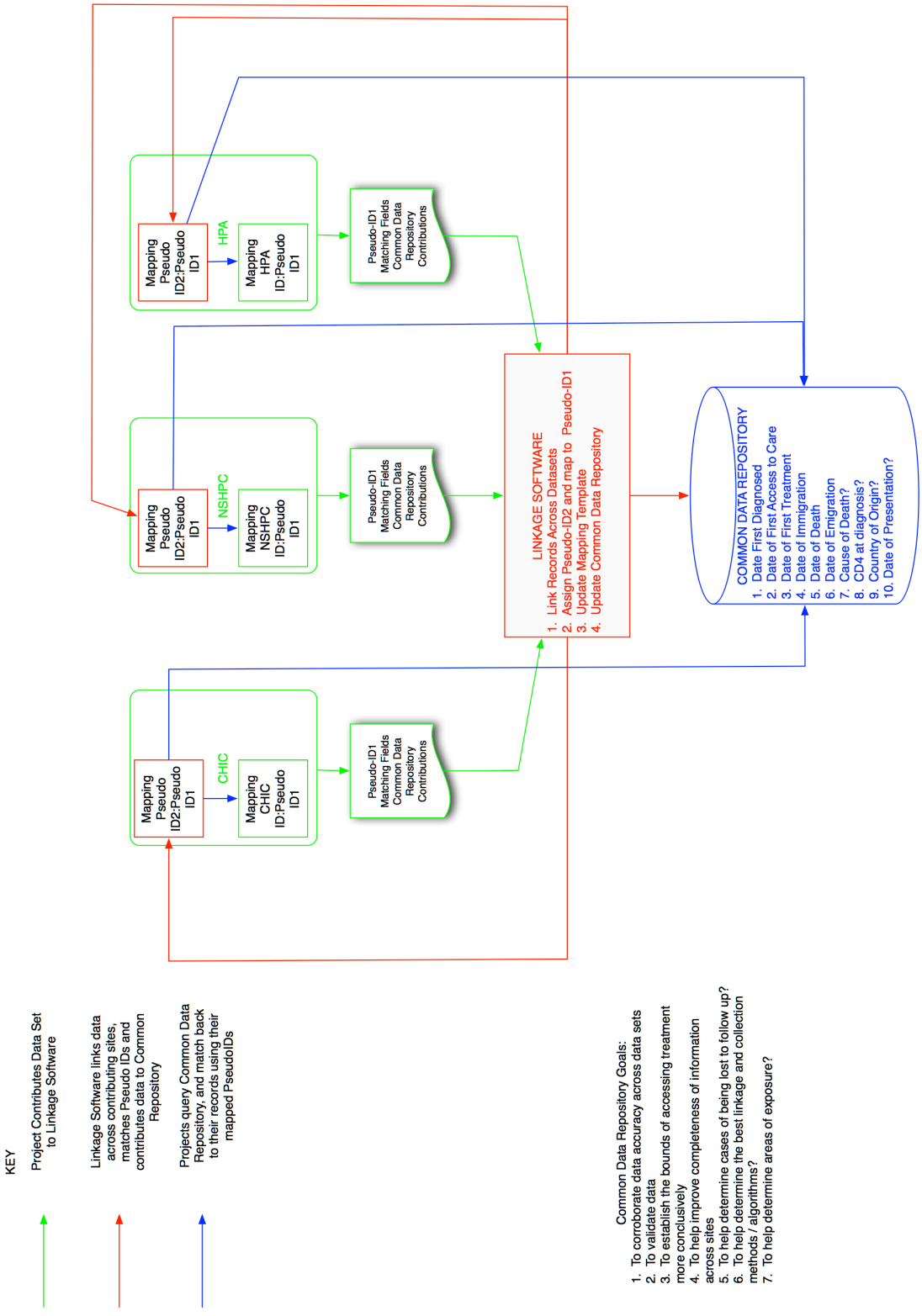


Figure 20: Specification of Linkage and Data Flows for DHICE Common Data Repository

5.2.2. Stakeholder Engagement and Contributions to the Research Work

The DHICE Initiative case study provided the author with an opportunity to understand EHR reuse to support research in the area of genitourinary care, which has considerable information governance requirements and expectations given the sensitivity of the information. The collaborative research between the studies with their separate databases also allowed a review of how separately developed information governance frameworks had to be adapted to support collaboration. By reviewing the information security policies and data sharing agreements, the author was able to get an insight into the layout and structure of existing information governance policy documentation that had been independently developed for research studies that were running, review the kinds of information that these documents contained and develop a common template for information security policy and practice, which was incorporated into the requirements analysis for the knowledge model.

By interviewing the researchers and data managers within the cohort studies, the author was able to gather a better understanding of information management and protection within an existing governance framework. This review of existing data holdings, practices and existing linkages gave the author first hand experience in developing linkage techniques and managing the implementation of new linkages, supported by the privacy enhancing methods developed by Kuzu et al. (Kuzu et al., 2013). This helped to develop the author's understanding around the governance requirements of establishing a new set of record linkages between the studies, integrate them within the existing policy frameworks and help to inform the requirements and development of the knowledge model for a collaborative working environment.

The DHICE Initiative also provided an opportunity to engage with HIV community advocate groups to help them understand the governance framework that existed to protect their information. The author was invited to give a talk at a UK Community Advisory Board meeting to discuss information security of patient records to help them understand how information is protected, and answer their specific questions:

1. What happens to information collected about me?
2. Who has access to it and how is it used?
3. How is it protected?

This allowed the author to articulate an overview of information governance for public engagement based upon the development of the common policy framework. This interaction also helped the author to appreciate the concerns of a treatment advocate group that could be fed in to the requirements analysis that is presented in Chapter 6. The presentation is available in Appendix 4.

5.3. Farr Institute for Health Informatics Research

The Farr Institute was introduced in section 4.2. Using this initiative as a case study has allowed the author to consider information governance requirements across a uniquely varied series of research projects and data handling measures. The Institute is spread across four centres in the UK, each with their own existing research data holdings and projects. These include larger population scale, epidemiological studies that contain centralised data repositories with their own existing information governance frameworks, as well as the more decentralised, federated examples as seen in the DHICE cohorts. In each case, the four Farr centres each have their own independent initiatives to develop secure data processing infrastructures and data safe havens in compliance with not only the Health and Social Care Information Centre IG Toolkit compliance, but also in the case of the Scottish Farr Institute, in compliance with the Scottish Health Informatics Programme's information governance expectations as well as the Welsh CIPHER example.

5.3.1. Farr Institute Information Governance Development

The author has worked as a member of the Innovative Governance group within the Farr Institute (Farr Institute for Health Informatics Research, 2014), which aims to develop a common understanding across the four Farr Centres of information governance requirements and procedures, for example a common template for data sharing agreements between Farr centres and a consistent set of standards and requirements for data safe haven services. The author has been

involved in the development of the London Farr centre's Data Safe Haven service in the capacity of a member of the project board and of the project executive of the de-identification and Research Data Indexing Service (RDIS).

The data safe haven has been developed as a service at University College London and has been developed to handle identifiable data. It provided the storage and software capabilities within a secure environment and offered a service to help research projects achieve IG Toolkit level 2 compliance. Having achieved IG Toolkit level 2 compliance itself, the Data Safe Haven (described in section 4.3) was developed according to the Information Governance Review recommendations and achieved ISO 27001 certification. By December 2014 it will have deployed the Research Data Indexing Service, which has been designed to de-identify identifiable records held within the safe haven and export it in accordance with data sharing agreements and the other information governance policies that research projects using the safe haven must adhere to. This service will help to keep a record of which records are available within the safe haven environment, link them when appropriate and strip the identifiers when data is shared with appropriate approvals. By working in the capacity of a data safe haven project board member and project executive for the RDIS, the author has been able to develop an understanding of the information governance policy framework needed to achieve IG Toolkit level 2 compliance and ISO 27001 certification. It has also helped the author to understand how researchers process information and their support needs for information governance of their working practice in addition to the regulatory framework within which a data safe haven service must operate. The design of a de-identification and data indexing service has also allowed the author to gain more experience and understanding of the various de-identification strategies and their use for providing and means of indexing the records that are available within the data safe haven.

Figure 21 shows the data flows from a variety of information sources into a data safe haven at the Farr London centre. Information sets and other assets are shown as "pots of gold" where the goal of the Farr Institute is to be able to share de-identified information across other sites using de-identification strategies.

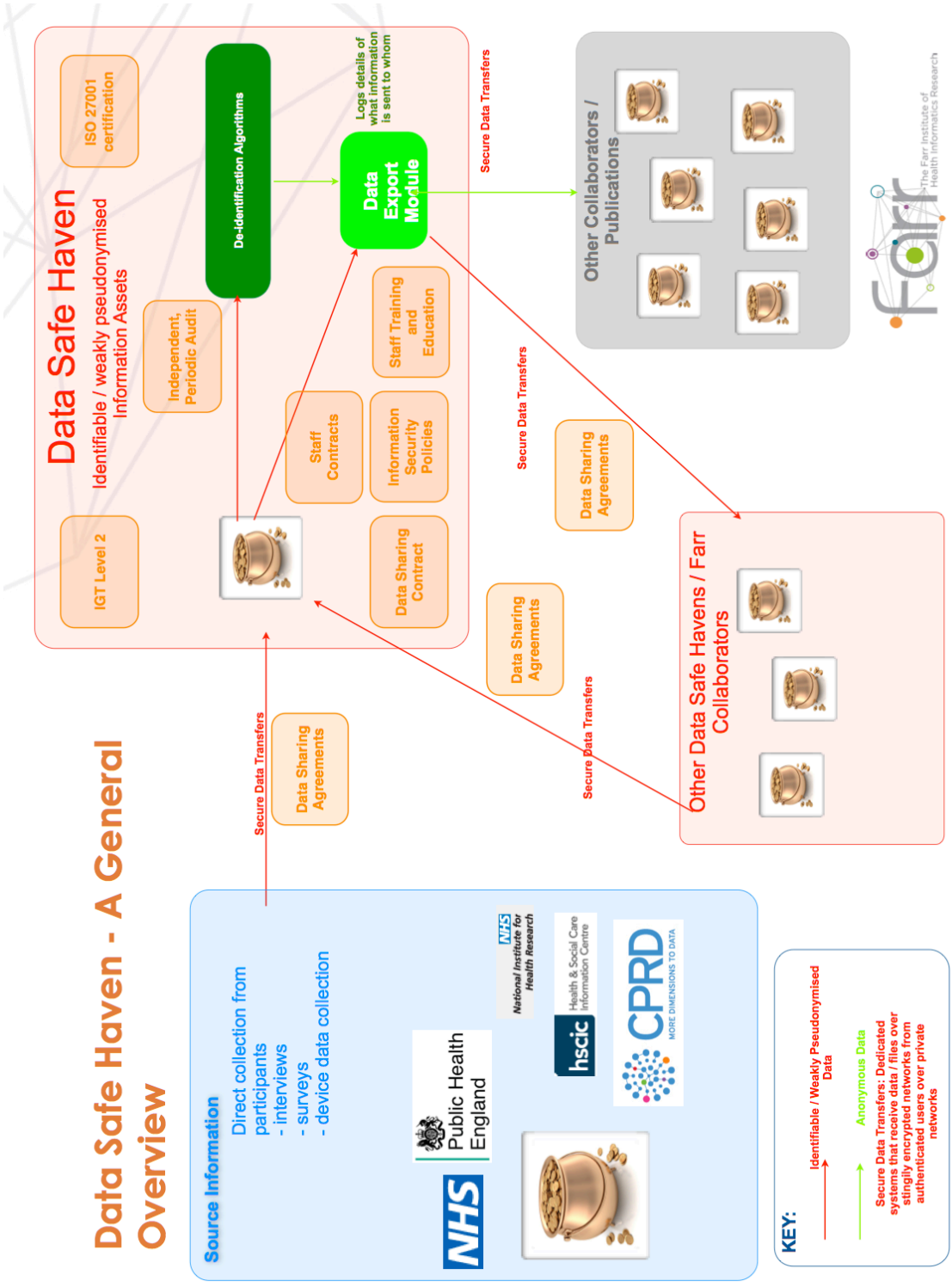


Figure 21: London Farr Centre Data Safe Haven and Governance Framework

Red lines show the flow of identifiable information and green lines de-identified information. The orange boxes show examples of the information governance control structures that are required and in place for the Farr London data safe

haven. These include certification through the ISO 27001 process and compliance with the information governance toolkit. Other examples include data sharing agreements that need to be set up between source suppliers of EHRs and recipients such as projects within the Farr Institute, as well as between the different Farr centres and other research projects. They also include reference to staff contracts, staff training and wider user engagement as well as the information security policies and data sharing contracts that are currently required for HSCIC Data Safe Haven accreditation.

5.3.2. Farr Institute Stakeholder Engagement and Information Governance Knowledge Model Development

The Farr case study has provided the opportunity to observe how the information governance requirements identified in CLEF and DHICE have continued to develop and evolve, including further refinement of security policies, data sharing agreements and the introduction of data sharing contracts. These in turn have been supplemented by a much keener focus on engaging with people who are involved in the handling of the EHRs in the research context, as well as those who are responsible for running the information services such as a data safe haven. This provided an opportunity to augment the requirements to allow for changes in legislation, standards and thereby information governance structures, as well as the need to support user engagement, education and training as another use case beyond advisory on how to handle information assets in a given context of use.

Stakeholder engagement has focused more on researcher needs within a context of use that includes multiple chronic condition and acute interventions. The potential linkages with social care and administrative data have provided new ambitions about the potential for data sharing, but also wider concerns from the research and other communities about the risks to the individuals about whom this information has been collected, their rights to consent and opt out as well as the higher risk of re-identification given the richer data sets that could be developed. This has helped to enrich the assurance requirements for the knowledge model and associated tooling, as well as the need to develop models that are generalisable for more than satisfying information security policy needs,

but also the educational elements as described above, data sharing agreements and contracts, as well as the assurances about how information security policies are developed, updated and used.

5.4. Electronic Health Records for Clinical Research (EHR4CR)

This project was introduced in section 4.2 and was used as a case study because, in contrast to the Farr Institute work, CLEF and DHICE initiative, it has developed an information framework that manages EHRs to support clinical trials of new drug treatments across participating hospitals in Europe. The project operated around three scenarios: the first was to assess the feasibility of running a particular trial by running distributed queries across all participating sites against de-identified records to see how many potential participants matched the clinical criteria specified in the query; the second was to manage the recruitment of participants into trials; the third was to manage the trials participation, integrate the trials data back with the participant's EHRs and to report on adverse affects.

5.4.1. EHR4CR Information Governance Development

Whilst working on the information governance work packages of this project, the author was able to review the requirements, access needs and information flows that depended on regulation and EHR databases from other European countries. The governance requirements were also different for each of the scenarios, where the first permitted the querying of only anonymised, clinical data, whereas the second required contact with potential participants through their treating physicians and the capture and reuse of identifying data. The third scenario required access and contribution to the EHRs as they managed the clinical trials participation.

Figure 22 shows the information flow for the first scenario. The components for this scenario involved the development of a query workbench for the development of feasibility queries. A component called an orchestrator would select the appropriate endpoint to send the queries to, whist these endpoints would broker access to a series of independent clinical data warehouses that contained anonymised data derived from EHRs and be managed within the

participating hospital site IT infrastructure and network in order to satisfy the legal requirements and access only anonymised data without consent.

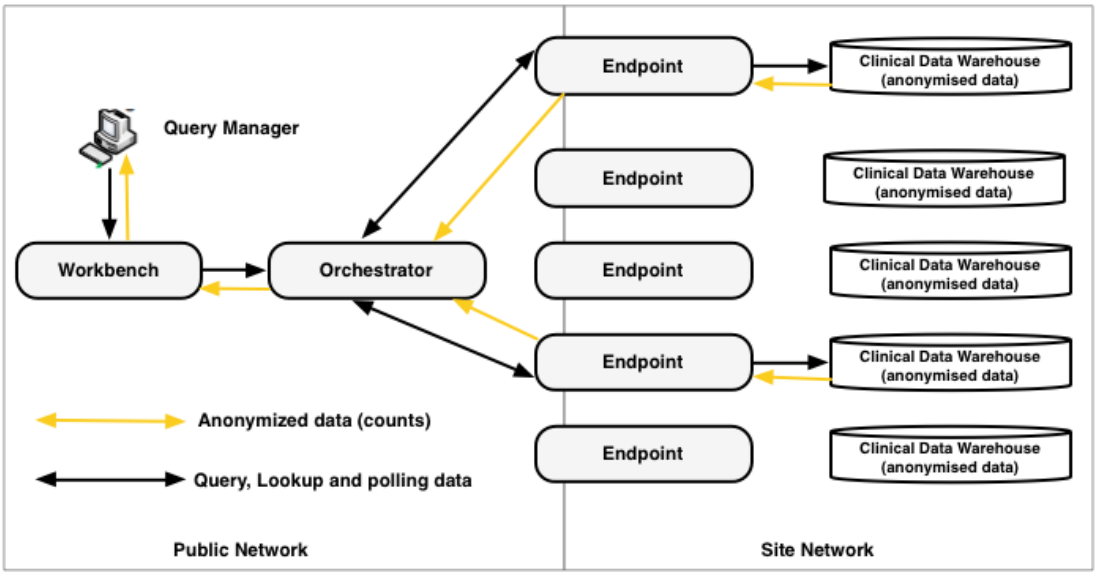


Figure 22: EHR4CR Scenario 1 Information Flow

Figure 23 shows the platform infrastructure for the second and third scenarios in addition to the first, where identifiable data is also being processed. This opened the use of the system up to other users, including study managers, data managers, treating physicians and trials investigators. It also developed other infrastructures that would be used in conjunction with live EHR systems containing identifiable data and not derived clinical data. The diagram shows the re-identification process to protect the identity of the participants from all but the treating physician and study investigator, as well as separating out anonymised data that could be accessed outside of the clinical site network on the EHR4CR Central Platform, which is accessible by study managers for the purposes of feasibility assessment and recruitment, managing aggregated information that is not readily identifiable. The author worked on the deliverables that handled the legal interpretations and requirements for the platform development, defined the technical security implementation requirements and developed the basis for the additional information governance policies that would need to be developed.

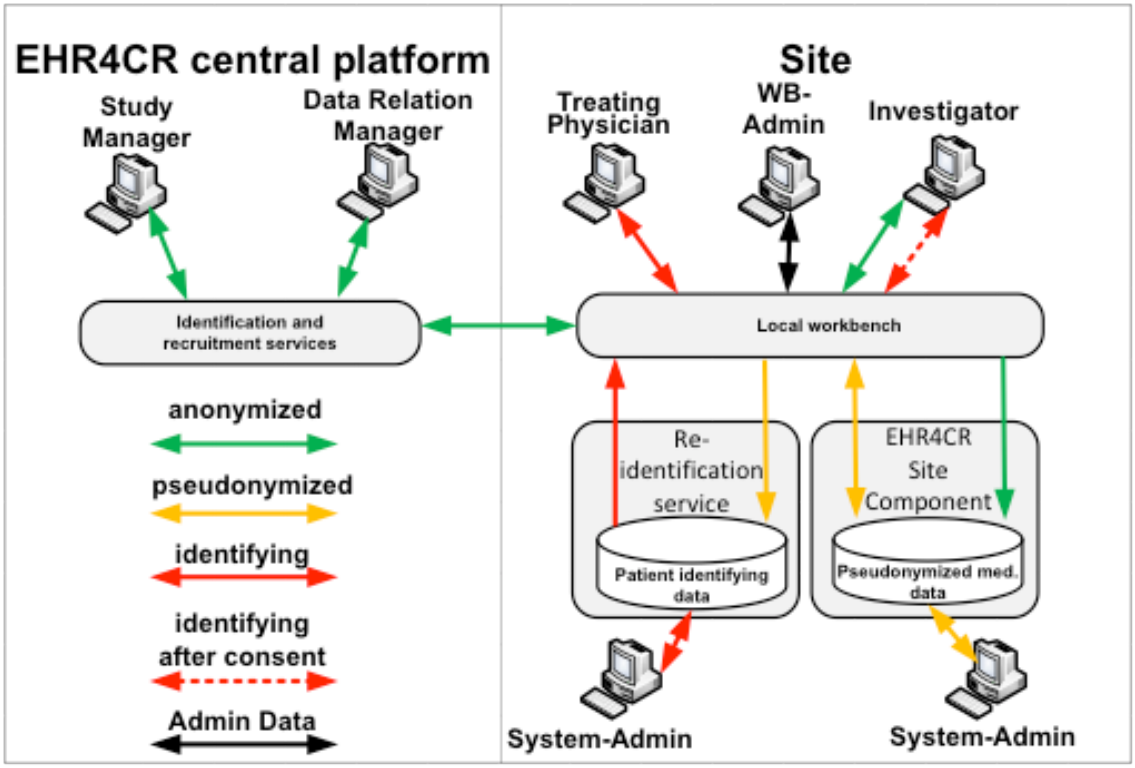


Figure 23: EHR4CR First, Second and Third Scenario Platform Infrastructure

5.4.2. EHR4CR Stakeholder Engagement

The information and workflows for EHR4CR illustrated in Figure 22 and Figure 23 demonstrate the breadth of stakeholders that were involved in this project, as well as the separation of identifying and de-identified data across the hospital site and wider EHR4CR consortium platform. This had specific challenges for the information governance work packages, which had to adopt a strategy of developing a Code of Practice that would apply to the whole EHR4CR group or any potential users of the system, whilst also developing standard operating procedure that would be applicable at the individual clinical site and study level for running the workbench components within that infrastructure. An illustration of the governance framework applicable in the UK is shown in Figure 24, providing an overview of which information governance control mechanisms would be put into effect to protect this particular flow of information.

Working on the EHR4CR project has provided the author an opportunity to further enhance the requirements gathering for the information knowledge model,

its implementation and use by including the secondary use of EHR information around supporting clinical trials.

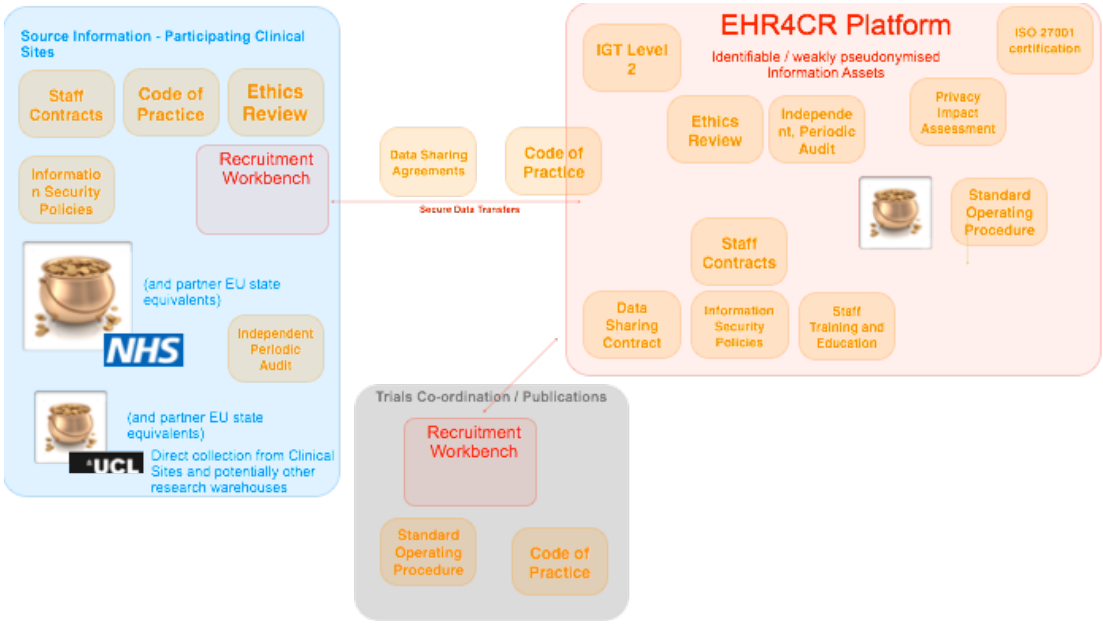


Figure 24: Information Governance Framework for EHR4CR - a UK Example

This represents a series of additional stakeholders, including representatives from the pharmaceutical industry, sponsors of research trials including UCL's Joint Research Office and a researchers working on a clinical trials and not purely research or surveillance studies. The concerns of these stakeholders were very much focussed on preventing the identification of the trial participants as with the other case studies, though in this case there was a stronger concern about correlating the access to information with consent, and the need to ensure that identifiable data would not be processed without the express consent of individuals if they opted to participate in a trial. Additionally, the pharmaceutical industry partners wanted to ensure that that their competitors would not be able to see the queries that were being run to protect their commercial interests. This had an additional requirement on the protection measures and added limitations to how auditing of platform use could be conducted in case their query was recorded in the logs, viewed by an unauthorised individual and thus undermining their commercial confidence.

This has helped to further enrich the author's understanding of EHR reuse, as well as augment the requirements for the knowledge model and tooling using a clinical trials specific information governance framework and input from stakeholders in the clinical trials context. By working on the code of practice and standard operating procedures, the author was able to examine the kinds of information held within these documents, which helped to further refine the knowledge model requirements and potential use cases for an information policy tool. This was particularly useful given that the information governance expectations would have to operate within several different EU jurisdictions and also needed to be consistently understood across the different users of the system, within the context of their home jurisdiction.

5.4.3. Integration with the European Medical Information Framework

In addition to EHR4CR, the EMIF project introduced in section 4.2 has provided a recent example of developing a common infrastructure across Europe for the collection of data relating to dementia and diabetes care to support research into these areas. EMIF is developing a similar infrastructure to the Farr Institute across Europe, which includes a common governance framework. This project has provided an opportunity for the EHR4CR governance team to adapt the Code of Practice to support this framework as well, illustrating how governance documentation like codes of practice need to have an inclusive scope. The Code of Practice has therefore been adapted to be inclusive of most of the IMI projects, which promote collaboration between academia, clinical partners and industry. This has provided the author with an example of how policy documents must be adapted to support a common framework, as well as be developed for a wide range of people who are involved in the different initiatives.

EMIF itself provided an opportunity to understand the potential use of a TransMART architecture for these projects, which are a cloud based system developed to encourage interaction and collaboration between industry partners (such as pharmaceutical companies) and research partners (Athey et al., 2013). The intention of the TransMARTs is to accelerate the translational benefits of research into clinical practice and provide a common framework for experts

around the world to contribute to the analysis of the available data. EMIF and other projects have raised some ethical, legal and security concerns about the TransMARTS given that they contain large amounts of anonymised data, which could be potentially identifying given the quantities of data used.

5.5. Case Studies Involving EHR System Development and Knowledge Model Design

This section describes the case studies that involved the development of EHR systems based upon a knowledge management framework as described in section 4.4. These case studies involved the author working on the development of archetype driven EHR servers for use in clinical practice and research. These case studies helped to illustrate the management of information governance in the context of clinical care and to develop an understanding of the requirements within that context, which was important for the clinical research context. It was during these projects and the ones described above that the Secutype model and Pattern Based knowledge management framework were developed.

5.5.1. The DebugIT Project

The Detecting and Eliminating Bacteria Using IT (DebugIT) project was a European Commission Seventh Framework project that researched the practical and legal requirements of sharing data relating to infectious diseases and antibiotic treatments across partner sites throughout Europe (Seventh Framework Programme, 2014). There was an emphasis on governance requirements in an attempt to establish legal sharing of data across European boundaries, where contributing and receiving countries might have different legislation governing the use of sensitive health care data. DebugIT provided an opportunity to explore different European governance requirements and working practice, leading to a means for reviewing how to harmonise sharing policy and permissions for collaborative European working, resulting in a higher level policy framework that would cover the project as a whole whilst refining specific details to each site across the European partners. This informed and guided the knowledge capture requirements to provide a consistent interpretation of governance stipulations

that would be acceptable to partners who were sharing data and understood by those receiving them.

DebugIT emphasised the importance of providing a consistent view of policy items, capturing legal requirements and matching expectations across a wider group of users. This was an important step to help internationalise the approach for capturing the details and guiding users, particularly given the promise of wider sharing across European boundaries. It was clear that basing the information management of a set of internationally recognised approaches would be key to implementing a usable system and that using the EN ISO 13606 knowledge model would be valuable for achieving this as the information modelling requirements became clearer.

5.5.2. The Dementia Register (DemReg)

This project involved the development of a registry for individuals diagnosed with dementia primarily for the purpose of clinical trials recruitment. Individuals diagnosed with dementia are sometimes deemed unable to grant consent for participation in clinical trials or for their data to be stored in such registries and it is left to their proxy, sometimes a relative, carer or medical practitioner to provide consent on their behalf. There were therefore more complex governance requirements to be identified in this case study. As the project extended nationally, the sharing of information was expected to require Section 251 exemption, as well as the running of a Privacy Impact Assessment (Information Commissioner's Office, 2009), which provided the opportunity to observe these processes being performed as part of this case study and further develop the analysis of the problem area accordingly.

The Register is provided by a state of the art implementation of an EN ISO 13606 compliant record server and a web application that is conformant with EN ISO 13606 part 4's Privacy Management and Access Control specifications. A case study of this project therefore offered an opportunity to explore and review the implementation of an existing health care record security management architecture, where a record specific to dementia care was established according

to an agreed model, and that model was used to generate the record editing and review screens within a medico-legally compliant record server architecture.

5.5.3. Heartbeat AC Development and Deployment:

The record server and web application architecture used for DemReg has been used to run an anticoagulation system for the monitoring of individuals who are treated with warfarin within the Whittington Health service. The service is currently used across North London hospitals and community healthcare providers, composed of GP surgeries and pharmacies. Heartbeat AC provided facilities to store clinical information, patient demographic details and details about clinic contacts. The system also provided a decision support feature that will review an individual's records and offer advice on the dosing of warfarin and an interval before an individual should next be seen based upon their International Normalised Range (INR) results obtained from blood tests. A significant feature that exemplifies a secondary information use scenario is the Clinical Governance, which provides statistics on numbers of patients seen, the percentages of INRs that remained within a specified range and how these are distributed across sites and clinic operators. The system is being adapted to a Stroke Prevention Service across North London and Hertfordshire.

The author worked on the deployment and configuration of Heartbeat AC, as well as the implementation for the system's specification. This provided an opportunity for the author to familiarise himself with the record and security architecture to a greater degree than the DemReg counterpart system because this system is now running in live clinical practice, which provided further opportunity to review governance requirements and data release flows for a series of users that have different data access requirements, be it for clinical governance, individual care or operator administration. The system has helped to expand the delivery of a newer, stroke prevention service. This has required clinical community member engagement, training and accreditation. The experience of deploying the system and helping to train users for the system provided valuable insight into how users responded to record keeping systems that are underpinned by a standardised information model and archetype constraint model.

5.5.4. Cortext Dementia Register

The Cortext Dementia Register was developed for a project to update and integrate a set of dementia research data from multiple research groups within the UCL Dementia Research Centre. During this work, the EHR Group at CHIME started to develop a new constraint model that would simplify the Archetype approach and allow for the expression of concepts in domains beyond healthcare record management, specifically the information governance domain. This new constraint model would be use to express clinical and information governance concepts and generate a web application in a more automated fashion than had been achieved using the second record server architecture described in sections 5.4 and 5.5.

The new knowledge model was called the Pattern, and is described in more detail in Chapter 7, along with the framework for generating web applications according to the specified knowledge models. Cortext was the first project that used this new approach, and the approach and its implementation were validated according to the requirements of the Cortext project as well as the exploratory methods used to develop the Information Governance model.

5.6. Concluding Remarks for Case Studies

The case studies helped to illustrate a number of key themes. They provided a more detailed understanding and realisation of the extent of the stakeholders involved with clinical research. They also helped to identify their roles, interests and responsibilities in clinical research. In some cases they emphasised the importance of developing an easily understood, approachable method for a range of people to understand their responsibilities when handling medical information. In other cases, they clarified the uncertainties and anxieties that public and patient representatives had and the importance of accountability, transparency and assurance for the handling of healthcare information as research was conducted. There were other stakeholders responsible for determining whether research should be permitted and for protecting that information, who helped to show that developing an understanding about data use and protection requirements needed

engagement, discussion and education as well as clear descriptions of the risks involved and how to mitigate against them.

A key finding from the case studies was the level of collaborative working that has been developing over the last decade for clinical research projects that make use of EHRs to underpin their work. These collaborations have developed into a series of consortia across national and European boundaries. The goal of these collaborations has been to help encourage a sharing of different expertise to answer clinical research questions, as well as to try and share data sets and build up a more complete picture of the health and increasingly social care experience of the UK and European populations. This helped to emphasise the new challenges to effective information governance, given that the collaborations would involve more people with different backgrounds and expertise, as well as more data sets that had different information governance requirements associated with them.

In addition to the trend towards collaborative working, another trend was the focus on integrating existing information governance frameworks and developing a combined policy framework between collaborating partners. This was an important observation because of the desired integration of existing information governance approaches and the development of fewer points of reference for understanding the expectations for working practice and behaviour. The need to simplify the process and availability of fewer, more detailed and consistently presented resources to guide behaviour and security software configuration needs was clear.

The case studies also provided a means to make direct observations of how information has been processed and queried to support clinical research questions. The author was introduced to paradigms such as the Data Safe Haven, which have been developed to provide a means to store identifiable data within an accredited framework, and TransMARTS, which have been designed to allow the querying of large amounts of de-identified data and pose their own ethical and security challenges. This meant that the technologies used to manage the information could be scrutinised, which allowed for an understanding to be developed about the use of technical security measures based upon the information governance policy frameworks. This also illustrated the issues around

interpreting those frameworks and putting the requirements into working practice. The use of knowledge management frameworks provided a means to assess the approach as a means to manage the information governance requirements, and to form a basis to develop a solution to manage the information governance needs within the collaborative research setting.

The next chapter provides the results of the learning that has come about as a result of the case studies, literature, legislative, standards and guideline reviews. These have provided a set of requirements for a knowledge management framework, which have been analysed to design and develop it. The approach and methods for the development are also provided and the implementation of the knowledge management framework for information governance is described in Chapter 7.

Chapter 6. Knowledge Management Framework Development

The legislation, policies, academic literature and case studies described in Chapters 3, 4 and 5 show that any processing of healthcare records is subject to a plethora of legislative and good practice requirements to protect sensitive information, and that meeting these requirements in a computable and scalable way is not straightforward. The requirements have been traditionally met using a combination of policy-based guidance to help people work according to safe practice and a series of software tools that can automate certain protection measures, both of which are informed by conducting a risk assessment and analysis to manage risks to the organisation. The software tools are not designed to communicate with one another, however, and the policy frameworks are sporadic and rely on narrative to convey the key requirements, which is prone to ambiguity, misinterpretation and may not even be read at all. The requirements of good working practice rely on the availability of clear, consistently understood information to provide knowledge about how to behave with healthcare records.

By reviewing the literature and conducting case study investigations, this work has been able to assess shortcomings and difficulties in working practice. The lack of a single "go-to" reference framework for policies is a key problem for informing people about how they should behave, and also encourages a lack of engagement and concern that behaviour is not as it should be. Web based applications have repeatedly been shown to inform and guide user behaviour through advice or decision support, often using a knowledge managed solution. The potential value of a knowledge model that would capture the requisite details for effective information governance to inform good practice is therefore self-evident.

The thesis of this research has proposed a knowledge management approach to tackle these problems. The approach capitalises on the dual modelling approach, defining a constraint model to effectively govern a reference or information model that is intended to represent information governance concepts and inform the development of an information governance management tool. These culminate in

the knowledge model that the thesis proposes, which the author has called the Security Archetype or *Secutype* and the development of a supporting infrastructure to provide the necessary information to help implement information governance and inform those handling EHRs or data items derived from those records. The aim of this tool is to clarify the requirements of information governance and security to the people who are handling sensitive data for research by providing them a single point of reference. The framework has been developed to encourage a consistent interpretation and understanding of the requirements so that user expertise can be supported and that they can behave according to the information governance expectations that are provided in the tool.

This Chapter describes the approach taken to develop the knowledge management solution. The approach used to develop the knowledge model and has followed the widely recognised Unified Software Development Process (USDP) as defined by Jacobson et al. (Jacobson et al., 1999), which proposes a five step, iterative process for developing software: requirements gathering, requirements analysis, software design, implementation and testing. This process uses the Universal Modelling Language (UML) for the notation of the design elements. The chapter commences with a section that summarises the main drivers for the development of the knowledge management framework derived from the literature review and investigations, from which a set of requirements has been distilled. These requirements are then analysed as a list of information templates, UML use cases and sequence diagrams, which help to inform the development of UML class diagrams and have been used to guide the implementation of the knowledge management framework described in Chapter 7.

6.1. Main Drivers Identified from Literature Review and Case Studies

The research work described in the previous chapters has illustrated the problems associated with the management of information governance requirements in handing healthcare information and sharing EHRs for care and research. This section summarises the primary requirements sources in terms of legislation, guidance and standards and the evidence-based problems identified from the

research work, which have been used to develop the requirements for the proposed knowledge management solution.

6.1.1. Key Legislative and Guideline Requirement Sources

A significant requirements source for developing the Knowledge Management Framework has been the ISO 27000 series of international standards on information security management, which focuses more on the practicalities of information security – the *who*, *where*, *when* and *how* of information governance, and in particular the ISO 27001:2013 standard, defining the requirements of information security. ISO 27002:2013 provides guidance on the implementation of the requirements to specific organisations and contexts, but 27001 has been a significant source of requirements for developing a knowledge management framework.

As to the *why*, the legal and ethical framework has provided a rich source of requirements for the knowledge management framework. In particular, the Human Rights Act, which defines the right to personal privacy across the UK, as well as the Common Law Duty of Confidentiality, which is the basis on which consent must be sought from individuals before any information held in confidence is released. These legislative instruments have been supported and emphasised by the Confidentiality Code of Practice, which emphasises the need for clinical practitioners and researchers to use identifying, confidential information only when necessary, to seek consent from the individual about whom the information has been recorded and to inform them of its use.

The *what* is data protection, represented by the Data Protection Act, which has emphasised the importance of specifying clearly use and purpose of personal identifiable information. Other legislation such as the NHS and Health and Social Care Acts has ramifications for the processing and treatment of personal healthcare information. The NHS Act defines the power for Secretary of State for Health to grant exemption to the duty of confidentiality under section 251 in England and Wales, described in section 3.1.2. The process and required information to make an application is a source of requirements to help develop a knowledge management framework. Applicants must be clear about how they

intend to process the information, the kinds of information they need to use and show that there is a substantial public interest.

The Health and Social Care Act 2012 provides a legal basis for the responsibility of processing of health and social information through the Health and Social Care Information Centre in England and Wales. This has helped to illustrate some of the proposed information flows that an information governance model should take into account. In addition to the legislation and the Confidentiality Code of Practice, the other guideline requirements sources that are now hosted by the HSCIC include various Department of Health Codes of Practice illustrated in Figure 7. These have helped to inform the development of information governance solutions in the case studies and illustrate the required use of the knowledge management framework itself, in addition to the knowledge model requirements. The Information Commissioner's Office templates for data sharing agreements have also provided a rich source of requirements for capturing information with regards to sharing information between organisations.

6.1.2. Observation of Issues with Current Practice

The literature reviews and case studies provided a very clear example of the issues that were evident with current practice and the motivations for this research work in proposing a knowledge management solution. These observations have fed into the requirement gathering and analysis for developing the knowledge management framework.

An important approach for developing information governance good practice and adherence is the consideration of risk to an endeavour, the organisation that is responsible for that endeavour and the stakeholders involved. Risk analysis and assessment are essential components in establishing protection requirements and developing countermeasures; an effective analysis depends on gathering details about assets, their vulnerabilities, threats and estimating the likelihood of a threat taking advantage of a particular vulnerability. The ability to establish metrics that determine whether risks should be protected against, commitment to information security and management priorities is regarded by legislation and standards as

essential for providing assurance to participants, funding bodies, governance bodies and other stakeholders in the research and clinical care domains.

Risk assessment and analysis help support the development of policy-based controls, which are critical to establishing how data users should behave in a given venture and organisation and for specifying responsibility and reference material. Policies, data sharing agreements and contracts, codes of practice and standard operating procedures are made up of a variety of documents, some for higher level human understanding, some for specific components like access control policy according to the technical specifications of the systems that are in place for a given organisation that manage access and privilege controls.

The process of defining security policies and a protection architecture is iterative and requires the capture, analysis and specification of a wealth of knowledge artefacts: these artefacts will in some cases need to be refined down to computable heuristics, or reused at different points in the analysis lifecycle, or whilst EHR systems are running and supporting research. Detailed knowledge about intent, practice, sharing and use of the data are core elements of security and governance controls that need to be asserted in a given use scenario. Some policy items can be automated in a running system environment to control data release and use, whereas others are specifically designed to establish the responsibilities and a code of practice for users.

Many policy items are applied manually and it is difficult to measure their effectiveness, particularly where data release, access and use are concerned; it is also sometimes the case that there is no evidence that these items were applied or adhered to. The technological means by which information can be protected are rudimentary and not easily configured. They cannot protect individual data items because it is a complex matter to configure them, then evolve a high level policy down to a computable one in the context of healthcare record systems. Mapping of the high level policy items to low level computable heuristics in a given system is also a manual process, and remains unsupported; there are a number of computable formalisms that have been developed to assert authentication and privilege management, but these are highly mathematical and formal, and do not

offer a means to comprehensively refine many policy items beyond access control and privilege management.

In addition to stipulations in a policy, working practice will also include day-to-day examples of control assertion inferred by data managers and handlers from policy recorded controls. A means of managing the wealth of security knowledge does not exist but is required to maintain a complete, reusable, shareable and auditable resource that will support not only the process of policy specification, but also the assertion of controls at the EHR server and computational level. Practical and technical implementations and research for assuring security have focussed on access control and privilege management as a mechanism to facilitate this: there is a common agreement among the literature that extensive authentication and authorisation services are needed, but there is little focus on how to protect or control the release of data post authorisation. It is nevertheless clear that stakeholders in the sharing of information for research and other secondary uses as well as providing care are concerned about the risks involved with these activities and need reassurance that what they are doing is consistent with working practice, their efforts can be assessed and they can show compliance with legal and good practice expectations. Researchers, sponsors of researchers and participants require that the basis for good practice is transparent, well managed, appropriately updated and adhered to.

The implementation of information security management currently follows an approach that requires human interpretation of policy and guideline narratives so that security tooling can be configured. Human interpretation requires refinement of policy items that are expressed at a high-level, human understandable level of abstraction to computable heuristics that a given software solution can interpret; this is currently a manual process, and the knowledge that is used to configure the tools is only persisted insofar as the tool requires. Where formal representations that can be used in a computational environment exist, they cover specific areas of control assertion and are not designed to interoperate: a consistent issue in the literature is that comprehensive security management facilities are lacking in the area of electronic data control and that many of these solutions are untested outside of a laboratory environment.

Policies are written with reference to existing technical solutions for controlling data retrieval, use and exposure, which focus entirely on granting access or managing privileges to access data resources. Organisations that hold health care data, however, follow a data release model when sharing data for primary uses, research and surveillance: a research or surveillance project would not be directly accessing an NHS database, for example, but the NHS centre will release data to the research project, according to the governance procedure as described. There is currently no method to comprehensively capture the knowledge about the governance and security constraints; this has the effect that systems are not configured consistently and the potential to more finely and automatically control data release without the need to resort to manual approaches has not been realised.

These observations have been made through the literature, legislative and security software reviews and many interviews, in addition to the practical experience of implementing security controls and working with clinical information during the case studies. Knowledge that is gathered during the establishment of the information security policy and implementation of security controls is extensive, but it is usually maintained in a narrative format as part of information security policies, tacit domain knowledge, accepted working practice or risk assessments: essentially, in a sporadic and inconsistent fashion.

6.1.3. Ethical Review and Research Funder Requirements Sources

The justification for applying for exemption from the Common Law Duty of Confidentiality is consistent with the process for gaining ethical approval to do research through research institutions or the Health Research Authority, where a description is needed of what the research is intending to achieve and how it will be conducted. Ethics reviews are used as a means to assure that clinically sound, statistically valid and safe clinical research is being performed and that participants will not be harmed. Apart from the safety to individuals, from a privacy protection point of view they are primarily concerned with the use of biological samples, but have had to adapt their processes to consider research projects oriented around potentially identifying, sensitive but scientifically and

clinically valid information. This results in further terms of use and stipulations being imposed on each such research project, including protection against identification risks and destruction of the data used at the end of a given project. Institutional governance is designed to guide researchers as to appropriate behaviour with participants when engaged in medical research by interpreting legal requirements and establishing a need for ethical review. Funding bodies may also have an ethical review requirement for research projects.

Funding councils have also asked for data management plans to be provided with research grant applications. These involve a description of how data will be held, managed and curated as well as how suitability for sharing will be determined, access will be governed and discovery of data resources will be possible. These elements reflect parts of the guidelines of the ISO 27001 standard, and show the kinds of detail that a knowledge model will need to hold.

6.1.4. Expectations for Greater Accountability and Transparency

The literature review and case studies have shown that there are anxieties about inappropriate use of this data for surveillance and criminalisation, particularly where the collection of routine health and social care information provides a resource for government surveillance. The research and clinical communities must be prepared for greater transparency, accountability and the ability to offer firm assurances about how information is handled not only when delivering care, but also for conducting research, which is becoming increasingly critical to maintaining public and stakeholder trust in the information sharing agenda.

A means of providing the public with some reassurance, honouring guidelines, working practice and ethics committee stipulations and complying with the Data Protection Act principles is the use of de-identification. Existing requirements for protecting information focus prominently on limiting the possibility of identifying patients, particularly if data is shared for research or surveillance. The ramifications are that certain items of information that have research significance (for example dates of birth, postcodes) need special protection mechanisms to protect the individual from identification. A number of proposals have been offered to try to limit the risk of identification of individuals from data when used

in research without compromising the research goals and data integrity when accessed. These are represented by work on de-identification and are implemented by anonymisation and pseudonymisation, but as indicated in section 3.1.1 are continually shown to be unsuccessful in providing anonymity.

6.2. Information Flows for Care and Research

The drivers for the development of the knowledge management solution are summarised in Figure 25 and Figure 26 below, which illustrate a "before and after" view of information flows when applying the proposed Secutype supported approach. These diagrams were developed by the author and presented at the Medical Informatics Europe conference in 2008 (Lea et al., 2008) to describe the current information flow and lifecycle of information and the protection measures that were in place and manually applied as shown in Figure 25. This diagram represents a series of information governance measures that would be applied as the EHR is captured at the point of care, held in a shareable format using Archetypes, shared with research and used to answer clinical queries. The proposed flow of information when applying the Secutype approach is demonstrated in Figure 26. Here, the Secutype holds the details of the information governance measures and applies them automatically in a consistent fashion using a specified framework. The MIE paper summarised the problem area and proposed Secutype solution and Figure 25 and Figure 26 illustrate the proposed approach in the context of sharing EHR information. The following sections expand upon this higher-level description and present the analysis of the research area and specification of the knowledge management framework.

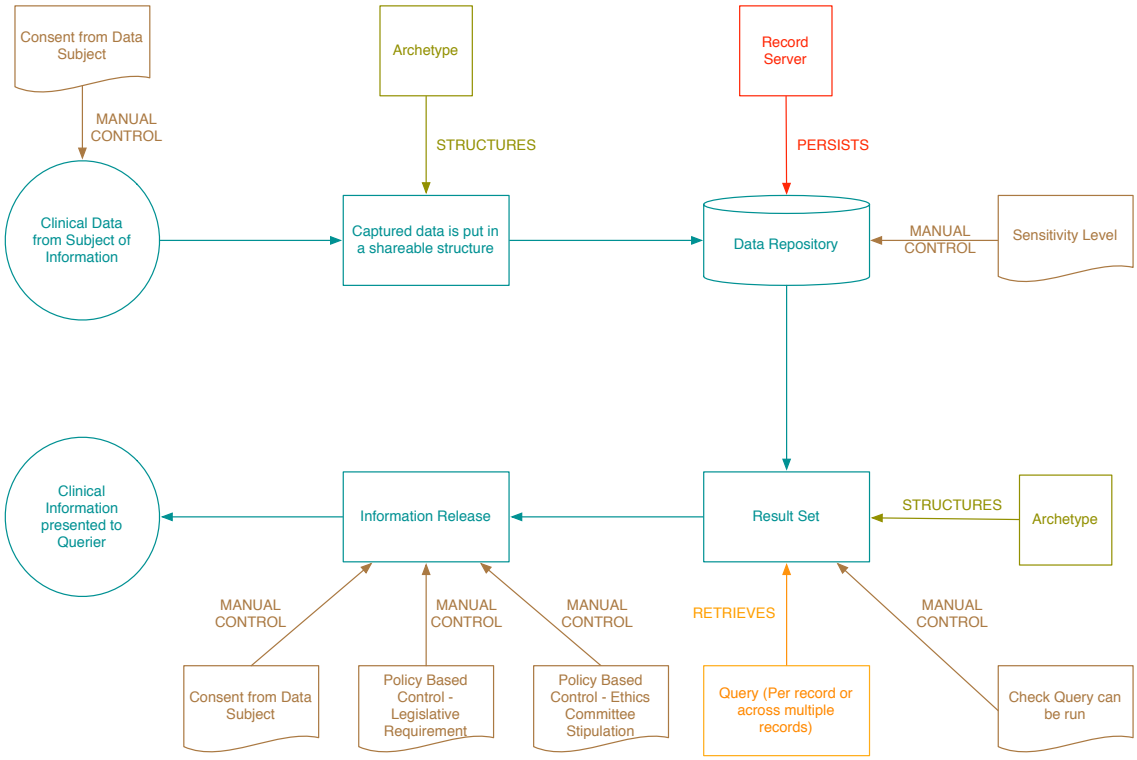


Figure 25: Flow of EHRs Pre Secutypes

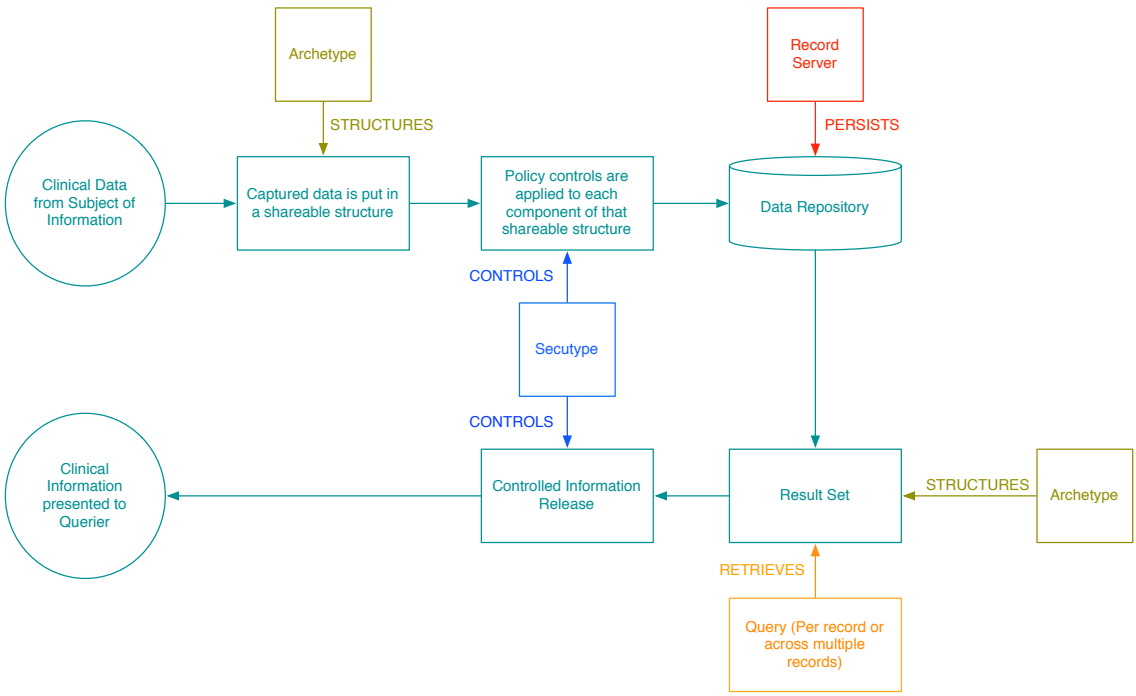


Figure 26: Flow of Information Using Secutype Approach to Assist Information Governance Requirements

6.3. Requirements for Information Governance Knowledge Management

The literature review and case studies established a rich source of requirements to help specify the knowledge models, the systems that would be needed to develop them and their use in practice. This section provides these requirements, upon which the Secutype has been designed, with reference to the requirement sources and how they have been used to specify them. As described in Chapters 3 and 4, the sources themselves include legislation, international standards for information security management and electronic Healthcare Record Communication, professional guidelines and codes of practice in addition to learning from the literature review and observations made during the case studies as described in Chapter 5.

A review of the requirements sources reveals a relationship between them, which is neither explicitly stated nor clearly established, as discussed in the earlier chapters of this work. This relationship resembles a hierarchy, particularly when assessing them as requirements sources for the knowledge managed solution. This hierarchy tends to start with legislative instruments, which set out the prescriptive expectations for protecting all stakeholders when handling healthcare information. These legislative instruments are refined down to more practical steps by guidelines from bodies such as the Information Commissioner's Office, the guidance for which is further developed and explained in practical terms by international standards for information management. These in turn are further interpreted by codes of practice specific to the type of organisations they are expected to operate in.

To illustrate, the Data Protection Act establishes a set of eight of principles which are more like generalised, prescriptive controls or rules that must be implemented when handling data, as also noted by Laurie and Sethi (Laurie and Sethi, 2013). For instance, the first principle requires that data uses are fair and lawful, whilst the seventh requires that appropriate measures be taken against unauthorised or unlawful uses of data. Another example is the Human Rights Act, which establishes in law a right to respect for private life that must be upheld and

any state interference must be justified on finite grounds and as necessary and proportionate. The Health and Social Care Act 2012 also specifies that the Health and Social Care Information Centre must process information with due regard to privacy. The Common Law Duty of Confidentiality requires that permission be sought from a subject of information before information that they have shared in confidence is shared more widely with others outside that relationship of trust unless that confidential information needs to be used in the public interest. The NHS Act 2006 allows the Secretary of State for Health to make legislation that sets this common law duty aside.

Taking the Data Protection Act Principles, the next level down in the hierarchy includes guidelines to help illustrate what they mean in more detail as general guidance. For example, the Information Commissioner's Office identifies the seventh principle as prescribing the need to process information securely. This guidance is itself further defined by a next level in the hierarchy of requirements sources, specifically information security standards like the ISO 27000 series for managing information security, where the meaning of security in more practical terms is specified in greater detail, defining the need for organisations to understand their own working practices, get managerial support and assess risks so that control mechanisms can be developed and defined. ISO 27001 defines the general requirements for information security management, whilst ISO 27002 describes the general code of practice for developing and managing controls for any organisation. Further contextual detail is provided by a standard for the implementation of codes of practice in healthcare settings in ISO 27799.

The final level of requirements sources includes professional guidelines and codes of practice within the context of healthcare information management, notably the Information Security Management and Confidentiality Codes of Practice, which help to provide a more tailored interpretation of some or all of these other requirements sources higher in the hierarchy. For example, the Confidentiality Code of Practice, though providing specific controls for organisations handling personal, confidential information also establishes principles around engagement with patients like that of "no surprises" for when patients are engaged about uses of their information. This code of practice helps to

provide a more discretionary set of guidelines and principles, emphasising the importance for members of an organisation to take personal responsibility when handling information and considering legal compliance with the Common Law Duty of Confidence, referring to data protection legislation as well as claiming that the Human Rights Act requirements for privacy will be met if data protection and confidentiality requirements are upheld.

The standards and guidelines lower in this hierarchy nevertheless refer back to the legislative instruments: the ISO 27000 standards, like all international standards, cannot mandate that the law be broken in countries adopting the standards but do cite legal restrictions along with risk assessment and organisational principles as requirements sources for developing information security management solutions. The NHS Information Security Management Code of practice refers back to ISO standards for information security management and the Confidentiality Code of Practice emphasises the need to handle information securely but relies on knowledge of information security practicalities as established by the standards. Additionally, the Data Protection Act becomes very specific in its definitions for responsible parties and stakeholders, including data subject, data controller, data processor and data recipient. The interpretation of these requirements sources in practice, where enactment of information governance requirements is arguably the most discretionary in the context of information processing, has been observed in the literature and the case studies (de Lusignan et al., 2007, Fernandez-Aleman et al., 2013, Manion et al., 2009, Al-Shahi Salman et al., 2014).

This interrelationship and varying degrees of specification across the requirements source hierarchy notwithstanding, the need for organisations to understand and interpret these requirements sources remains so that they can develop the information security controls and levels of good working practice that are expected, not only to achieve the protection of the information subjects and organisation, but also ensure that people responsible for handling the information are engaged, understand what is required of them and can work effectively. Observation of working practice during the case studies has helped to illustrate how difficult a task this is when bridging the gap between the guidelines,

standards and legal requirements and actual working practice. A core theme of this thesis is to help fill that gap so that interpretation can be aided and made more consistent, people governed by specific controls and policies are engaged and understand what is expected of them and that the wealth of requirements sources can be effectively and succinctly refined to meaningful guidelines and specification of controls.

The requirements below have been categorised according to four areas of use for the knowledge management solution. The first represents the general, higher level requirements for the knowledge management approach, its development and use in practice for information governance, which tend to relate to the prescriptive legislative instruments. The second area focuses on the Secutype knowledge model itself and the information that must be captured in order to support effective information governance, which rely to a much greater degree on the more specific illustrations of information security controls defined in the standards, other, higher level guidelines and observations of working practice in the case studies. The next area focuses on the creation, editing and management of Secutype instances through an editor. The fourth area focuses on the requirements for using instances of the knowledge model in practice to advise people and configure software on the handling of EHRs in clinical research.

The MoSCoW prioritisation scheme has been used to prioritise the requirements, where M stands for Must have, S stands for Should Have, C stands for Could Have and W stands for Want to have (but Won't yet). This work has focused on the implementation of M and S prioritised requirements, where the lower priority requirements have been proposed as further work. The requirements have informed further analyses, including use cases and sequence specifications. The knowledge model requirements have been further developed into a series of information templates, which in turn have been used to develop Class diagram specifications for the knowledge model itself. The further analyses are described in the subsequent sections of this chapter.

6.3.1. General requirements

Table 2 provides the high level requirements for the knowledge management solution, covering the specification of the Secutype, its development, management and the general principles for what it should represent, how it should be used and what the intended outcomes are. A discussion about the sources and provenance of these requirements follows. These higher level requirements are refined to more specific requirements in the following subsections, where a more detailed discussion of the sources and their provenance is provided.

Requirement Number	***The knowledge model driven solution shall***	***Priority***
GEN-1	capture specific items of information about a particular EHR processing project or use	M
GEN-2	gather information about the governance of clinical information use from multiple sources, including legislative instruments, stakeholder engagement, collaborating organisations and risk assessments and analyses	M
GEN-3	support decision making on whether to share information based on multiple sources, including legislative concerns, individual consent, general policy adherence and a need to know whether sharing information will be harmful or inadvertently identify a patient	M
GEN-4	support decisions based on interaction and collaboration at the level of human contribution, and refine the sources for decision making to a computable form for use in live software system environments.	M
GEN-5	store, represent and generate knowledge artefacts based upon risk assessments, information governance reviews and security policies that exist	M
GEN-6	capture and accurately represent the details that are provided in information security policies, data management plans, data sharing agreements and contracts	M
GEN-7	support the completeness of specifying policy documents and working practice guidelines	M
GEN-8	limit ambiguity and inconsistent interpretations of specified policy documents	M
GEN-9	help maintain the anonymity and limit the identifiability of participants when their information is used for research	M
GEN-10	maintain a complete record of information assets, information handlers, procedures and practice as required to effectively govern information use and working practice	M

GEN-11	keep a record of user interactions with policy documents to support a complete audit trail and provide appropriate assurance to data subjects, funders or other stakeholders	M
GEN-12	provide a knowledge managed solution in a software tool accessible by any user responsible for developing information governance solutions or governed by an information security policy and / or guidance for working practice	M
GEN-13	manage the update of policy elements and the knowledge model itself in line with changes to legislation, guidelines and standards	M

Table 2: High Level Knowledge Model Requirements

Requirement GEN-1 frames the overall purpose of the knowledge model driven solution, defining its scope and primary function in accordance with the thesis of this research. In general terms, this requirement has come from the review of the requirements sources described in Chapter 3 and observations made during the case studies, particularly CLEF and DHICE, where part of the work involved the development of information governance frameworks. The forthcoming discussion of the lower level requirements refers to their sources and provenance in more detail, though a primary source for this requirement is ISO 27001:2013 section 7.5, where the documentation and its availability about the organisation and information processing is identified as key to effective information security management, identified in more detail in ISO 27002:2013.

Requirement GEN-2 is also based on these sources. The legislative, standards based, good practice and professional guidance sources themselves recognise that there are many different sources for developing information governance of clinical information, and this has been made clear from observation of working practice and governance management in the case studies. The definitions of information governance provided in the introductory section of Chapter 3 emphasise the importance of legislation, particularly the Data Protection, NHS and Human Rights Acts, whilst the importance of international standards and independent accreditation are emphasised in the Information Governance Review, which in turn emphasise the importance of stakeholder engagement.

GEN-3 represents a key use of the solution, to aid in advising users on how to handle healthcare information and whether to share it, based upon a refinement of

the requirements sources to a meaningful set of directions. The process of ethical review is a pertinent source here, which considers the risk of harm to participants in research, including inadvertent re-identification of those participants as information is shared and linked. Re-identification could lead to inadvertent harm and is therefore a significant risk, and may change the legal basis for processing the information.

GEN-4 represents the need for the solution to tackle a key problem that the research is trying to resolve by maintaining consistency between the narrative specification of policy for human reading and its refinement to computable heuristics. This has been determined through the review of the literature and case study observations, as described in Chapter 5.

GEN-5 represents the need be able to generate meaningful knowledge artefacts generated from the identified sources. ISO 27001:2013 section 6 emphasises the importance of risk assessments and how they should be used to develop mitigation strategies and policy based controls, as well as periodic internal and external review. This importance is highlighted by the Information Commissioner's Office, which encourages the use of privacy impact assessments that can be used to inform a risk assessment and guidance on working practice. Ethical reviews are another form of risk assessment for not only the participants but also the institutions conducting the research. The requirement for the solution to generate guidance and informative knowledge artefacts has been made clear by the sources described in Chapter 3 and the case studies, particularly those that involved managing good practice as electronic healthcare records were being handled, where it was clear that there was no method to refine legislative, standards based and good practice guidelines to a specific use context and aid the development of policy and mitigation strategies.

GEN-6 takes several sources as its basis. The structure, importance and use of policies are emphasised in the ISO Series of standards, particularly ISO 27001:2013 section 5.2 and section 7. Data sharing agreement and contract templates from the Information Commissioner's Office and Health and Social Care Information Centre, whilst data management plans are proposed by the Medical Research Council and Wellcome Trust. It is clear from these sources that the

accurate capture and representation of these details is important for effectively managing information governance.

Requirements GEN-7 and GEN-8 share the same requirements sources and represent directly a proposed resolution to the problems identified in the thesis. These are intended to support GEN-9, which has been identified in the review of requirement sources as a key goal of information governance in the context of research using de-identified records. Ensuring that governance controls are complete, clear and unambiguously specified is required from findings in the literature as well as compliance with the first and seventh principle of the Data Protection Act where identifiable information use is concerned, but limiting re-identification risks is also key to these principles as well as those in the Confidentiality Code of Practice.

GEN-10 lays out the elements within an organisation about which information should be captured. The sources of this list is based upon the set of components that ISO 27001:2013 recommends are identified and used for developing an information security management system, managing risk assessments and definition of these elements within policy under Annex A. This supports meeting the expectations of the Data Protection Act principles and is also necessary for preparing other information governance guidelines, as discussed in the following subsections. This will also help to meet the requirements of the Information Governance Toolkit as specified in Figure 6, particularly requirement 11-210.

The source for GEN-11 is based primarily on requirements to show compliance, in particular in cases like P against Finland, where the defending hospital was unable to show how they were protecting the plaintiff's records, and the cases prosecuted and fines levied by the Information Commissioner's Office. Transparency has been identified as a key requirement for the processing of healthcare records as can be seen by the anxieties and reports on public trust and confidence in the handling of healthcare records referred to in section 3.3. The Information Governance Toolkit requirements 11- 205 and 11-206 are also requirement sources here.

GEN-12 is based upon the core need for a tool to encapsulate the information governance requirements within a knowledge managed solution as well as to help

provide a consistent view of those details to correctly advise users, where observations during the case studies emphasised the need for this tool. These are also established by section 7 of ISO 27001:2013, which focuses on support for organisation members to "…determine and provide the resources needed for the establishment, implementation, maintenance and continual improvement of the information security management system…" including determining competence (in line with requirement 11-112, 11-200, 11-300 and 11-400 from the Information Governance Toolkit).

GEN-13 is based upon the need for periodic review of policy in practice as specified in ISO 27001:2013 section 10, NHS working practice guidelines and the Information Commissioner's Office. This may occur if there are changes to working practice, information use requirements and availability of resources. The update to the knowledge model itself may also be necessitated by changes to legislation and other requirements sources, as well as recognised good practice, as observed during the case studies.

6.3.2. Knowledge Model Requirements

The Secutype must model the details that are needed to inform the development of policies, sharing agreements and good practice guidelines. The requirements for the Secutype are provided in this section. Across the requirements sources described in sections 3.1 and 3.2, a consistent theme emerged with regards to the scope of information governance, particularly from the ISO 27001 standards for information security management, which advocate the identification of the context of a particular organisation to "…determine external and internal issues that are relevant to its purpose and that affect its ability to achieve the intended outcome(s) of its information security management system…" (British Standards Institute, 2013b)This would depend upon the context within which healthcare information was being used, as it related directly to the activities that were being performed with the data, the people and software involved, the information assets being used, be they discrete data items derived from EHRs, entire databases, servers or USB keys and the legal bases which permitted the processing of the information in the first place.

Any of these details could also be gathered as part of a risk assessment process as well as feed into policy specification. The experience of working on the case studies described in Chapter 5 also showed that a use context should represent more than an entire organisation: information governance had to be developed for specific research departments, projects and other collaborations that occur within an organisation in addition to the entire organisation itself. The specification of the use context should therefore include a means to specify the scope of that context. Table 3 provides the requirements for the use context, which would contain the other core details needed to inform effective governance. The items contained within the use context are listed in more detail in Tables 4 – 7.

Requirement Number	***The knowledge model shall***	***Priority***
COU-1	faithfully and accurately capture details about the type of context where information assets are being used, be it a research institution, department, research or other secondary use project	M
COU-2	capture details about the people involved in a particular use context, their roles, responsibilities and identification details	M
COU-3	gather details about information assets that are used within a particular context of use, including any unique identifiers, the type of asset (be it software or hardware, a database or USB key, for example)	M
COU-4	contain details about the use and sharing of EHRs within the use context and the purposes for the use of information, including details obtained during a risk assessment	M
COU-5	contain details of all security controls and protection actions that are available within the use context	M

Table 3: Use Context Requirements

Sources for COU-1 establish the need to understand the context and its scope. They include the Data Protection Act, particularly the first, second and seventh principles, as well as the Information Governance Toolkit requirement 11-200. Another primary source for requirement COU-1 is ISO 27001:2013, particularly section 4, which focuses on the context of the organisation. Gathering these details is essential to developing a governance framework and helping to contextualise risk assessment and the development of policy items, as per sections 5 and 6 of ISO 27001:2013. There are a variety of different kinds of context of use as learned during the case study observations, where multiple organisations and research

groups were part of the CLEF, DHICE and EHR4CR projects and different purposes of use, users and information assets each with their own risk profiles and formed part of these contexts. The requirements for the knowledge model therefore include capacity to handle this variation.

COU-2, COU-3 and COU-4 are based on capturing details about elements within the context of use, including information assets, people involved and specific uses of information assets by those people, as established throughout ISO 27001:2013. Additionally, in the case of these, COU-1, and COU-4, section 7.5.1 of the standard provides an additional source for these requirements, where these items will determine the nature of the documentation required by the standard. COU-2 is the requirement for the knowledge model to represent the people involved in a particular context must also be represented in determining information governance needs for a particular enterprise or use. The Data Protection Act defines data subjects, controllers, processors and recipients, which may be organisations or people and where rights or responsibilities are conferred to these different role holders. ISO 27001:2013 is clear on the need to capture details about staff, their responsibilities and their contracts of employment. This is also true of third party vendors and suppliers, the details of which can be found under Annex A under sections A7 and A9. The Information Governance Toolkit requirements 11-110, 11-111 and 11-112 are also sources for this requirement. In combination with the roles and privilege management requirements established in EN ISO 13606 Part 4, the literature review on privilege management and development of the roles specification in CLEF as provided in Appendix 2. These are the sources for the more refined requirements for capturing these details are specified in Table 4.

Requirement Number	***The knowledge model shall***	***Priority***
PEI-1	faithfully and accurately capture identification details for people involved in a particular context of use	M
PEI-2	Specify the roles, responsibilities and job titles that each person holds	M
PEI-3	capture details about any affiliations that the person may hold with other organisations or use contexts	M

PEI-4	hold details of employment contracts or other documentation that are specific to the user	M
PEI-5	indicate when a user has been assigned or delegated a particular role, responsibility or function by another person temporarily or indefinitely	C

Table 4: People Involved Knowledge Model Requirements

The various requirements sources define responsibilities for specific individuals within a context of use, where these individuals may occupy roles in different organisations or other use contexts. These roles and responsibilities are defined in employment contracts, as illustrated by observation of the case studies. It was clear from the case studies that delegation of roles and responsibilities was present in the modus operandi and has been included as requirement PEI-5, though provided a lower priority in this iteration for the purposes of focusing on advisory for data handling and the complexity of binding users within this framework, where this is discussed under further work.

Information processing in a given context involves the use of a variety of information items or assets. A core element of the ISO standards and risk assessment approaches refer to the identification of information assets and keeping a register of these assets. The primary sources for COU-3 are section A8 of Annex A in ISO 27001:2013 and these are expanded upon further in Table 5. Section 5.4 EN ISO 27799 provides further examples of the kinds of healthcare information that needs to be protected, emphasising the different protection measures that need to be employed. The case studies also provided evidence for these more detailed requirements, particularly for the eScience and data management projects like CLEF and DHICE, which showed that information assets could represent anything from a single row in a database table to a research computing platform that includes a number of servers, disk arrays and network infrastructure. These case studies also showed that data formats and structures offered further details on how information could be represented and protected.

Requirement Number	***The knowledge model shall***	***Priority***

IA-1	faithfully and accurately capture details about the kind of information asset that is used in a particular context of use, whether it is a single data item, a physical, portable storage medium, a server or a series of databases	M
IA-2	specify the storage medium for this information asset where applicable	M
IA-3	capture details that uniquely identify a particular information asset	M
IA-4	specify the format data where appropriate	M

Table 5: Information Asset Knowledge Model Requirements

COU-4 refers to the purposes, uses and sharing of information in a given context. The particulars of data use are important for showing compliance with the first principle of the Data Protection Act. These uses must be aligned with specific purposes, the details of which are required for ethical approval and management plans. These purposes must be in line with a basis in law, as per the second Data Protection Act Principle, which uses the purpose to establish the scope of uses to achieve those purposes. Defining acceptable uses is a key requirement from ISO 27001:2013 under Annex A section A.8.1.3 and it is important to have clear and accurate information about uses of the EHRs or any data derived from them in order to understand risk and guide safe and appropriate behaviour in given circumstances.

These form the basis of the more detailed requirements for capturing information about uses and purposes of using information as shown in Table 6, specifically requirements USE-1 to USE-6, which establish the need to store details about use, purpose and the legal bases upon which they proceed as well as the relationship between these three concepts. The good practice guidelines and requirements for gaining ethics committee approvals as detailed in sections 3.1.2 and 3.1.3 as well as the case study experiences have shown the importance with which defining purpose and legal bases for specific activities has been held by various stakeholders in conducting research, as well as the wealth of different legal bases that are available. It is clear from the guidelines on performing research, funding body expectations as well as stipulations in Chapter 2 of the Care Act 2014 that these purposes and uses rely on legal bases such as consent. Other bases

include the approvals that may need to be granted so that research can be performed, or if exemption from the Common Law Duty of Confidentiality is sought under section 251 of the NHS Act 2006 in England and Wales, or Caldicott Guardian approval in Scotland. These are the sources for requirement USE-3.

Additionally, The Confidentiality Code of Practice emphasises the importance of making clear when identifiable information is being used, informing the data subject about the use where possible and when consent should be gathered as a legal requirement. This is shown in the table of Key Questions of Confidentiality under paragraph 38 of the code, particularly the last three points that emphasise the importance of defining purpose and the need the need to justify the use of identifiable information, whether appropriate steps have been taken to inform patients and determining whether explicit consent is needed. Annex A2 further establishes the last two points, by representing informing patients with the principle of "no surprises." These are the sources for requirements USE-7, USE-8 and USE-9.

Requirement Number	***The knowledge model shall***	***Priority***
USE-1	faithfully and accurately capture details about any use or sharing of clinical information, specifically the activities that are being performed with that information	M
USE-2	faithfully and accurately capture details about the items of information that are being used, including what they represent, how they are stored, any physical hardware items that might hold them and any processing software that might be used.	M
USE-3	allow for the detailed specification of the legal basis for any use of that information, including pertinent details (i.e. if consent has been gathered from participants)	M
USE-4	store details of the specific approval stipulations for use of the released information	M
USE-5	faithfully and accurately specify the context of use within which the stipulations apply	M
USE-6	faithfully and accurately specify the purposes of using any clinical information	M
USE-7	faithfully and accurately specify whether identifiable information is being used and the permissions to do so	M
USE-8	faithfully and accurately specify whether patients should be informed about the sharing of their data and what effort has been made to inform them about this	M

USE-9	faithfully and accurately specify whether informed consent is legally required for data disclosure to be lawful	M

Table 6: EHR Data Uses, Purposes and Legal Basis Specification Requirements

The specification of protection measures and controls on how information assets are used is perhaps the most significant component in the information governance requirements because these represent how information governance is asserted. The seventh principle of the Data Protection Act establishes in law the need for security to prevent unlawful processing of personal data, where the Confidentiality Code of Practice provides examples of specific controls for securely working with confidential information within its Figure 5, discussing how to keep patient information secure. This is consistent with the definition of controls as per Annex A, Table A.1 of ISO 27001:2013, where generic controls are specified to meet a series of objectives as set out by information security risk treatment defined in section 6.1.3 of the standard. These controls are defined in more detail in ISO 27002:2013, the code of practice for information security controls. Each control listed in Annex A of ISO 27001:2013 is illustrated by further discussion and guidelines for implementation, and additional information where appropriate. This standards also identifies the requirements sources for specifying controls, including risk assessment, legal compliance and principles established by organisations as they process information to achieve their goals.

The case studies illustrated that asserting controls was reliant on correctly specifying those controls in the context of the information assets that would need to be protected, the activities that were being performed with those information assets and the people authorised to work with them. Particular types of information assets would have a control applied to them, for example the use of USB keys would be forbidden for any purpose within a particular use context. Alternatively, as shown by EN ISO 13606 part 4, individuals should have the ability to restrict access to particularly sensitive information to specific individuals, representing an entirely different execution of the control for that particular context of use. Risk assessment has been considered in the literature and good practice guidelines as a primary source of specification for developing the guidance on how to behave with specific assets, though this is often accomplished

without direct reference back to organisational goals and the requirements of the people involved with the processing, a trend that was also apparent in the literature and good practice guidelines where a harmonised view of the information governance details and how they interrelate had not been articulated.

Whilst the legislative instruments define the need for security and controls, the standards identify generic controls in a given context of use and guidelines further refine the broader constraints in the context of working with healthcare information, the practicalities require that specification of controls is clear, related to purposes, the users that they apply to and the kind of information that is being used and how precisely how it needs to be protected for a given activity. In the context of this work, the word *control* is therefore used to mean the mechanism by which the use, processing or sharing of an information asset is managed in accordance with the protection requirements established by the principles established by the governance of research, in a risk assessment and / or specified in policy.

The guidance and specification of controls must take these into account and binding these concepts together was an important requirement for the Secutype knowledge model to help achieve its proposed goal of harmonising a complete representation of the information governance that is expected, whilst also developing the information governance strategies and tools with a complete, accessible reference to the specification details. These requirements are described in 7, where a view has been taken on simplifying and refining the details so that they can be readily used and related to the other elements within a context of use as described earlier in this section.

Requirement Number	***The knowledge model shall***	***Priority***
CON-1	faithfully and accurately capture details about the controls that should be applied to specific information assets or sets of information assets based on policy, approvals and approved activities, legislation and consent	M
CON-2	specify control details for specific activities and / or information users in a particular use context	

CON-3	specify the control mechanism applied, for example forbidding the sharing of information on a CD-ROM through a postal service or a particular de-identification algorithm	M
CON-4	record details on what options or methods for the control are available, for example what de-identification algorithms exist and how de-identification should occur	M
CON-5	store details about when and under what circumstances controls should be applied	M
CON-6	record under what context of use and for which users and at what times controls should be enacted	M
CON-7	capture details as to why the control has been enacted: based on what policy item, approvals basis, legislation element or consent stipulation	M
CON-8	specify whether a user of an information asset should be informed that data set has had items removed for anonymisation or amended for pseudonymisation	C

Table 7: Control Knowledge Model Requirements

The next section provides the requirements for managing the Secutypes in terms of authoring, storage and update.

6.3.3. Update and Evolution of Knowledge Model and Stored Details

The legislation, standards and guidelines that form the basis of information governance have been subject to several changes in the UK and beyond, as described in sections 3.1 and 3.2. The Secutypes themselves may need to be updated according to any such changes, as described in requirement GEN-13. This has been handled by editing tools in the case of the Archetype as described in section 4.1, where it is clear that a collaboration between knowledge model authors is essential for gaining a common understanding and agreement on how clinical concepts should be modelled. A similar approach is needed for developing security concepts in order to support a shared development and common understanding between authors of the Secutypes themselves, particularly where changes in core legislation, standards and guidelines would require the interpretation of information governance domain experts.

Table 8 provides the requirements for developing and updating the knowledge model. These requirements relate to the need for an editing tool to support this, with collaborative working functions. It should be noted that UPD-3 is classified as

a *Should* priority because it is likely that most changes would need to be made to details contained within instances of the knowledge model, not the model itself. UPD-4 is a *Should* because these relates to editing of the Secutype and not the policy and guideline details that it represents. UPD-5 relates to the update and evolution process, following the stages of model development that relate to Archetype authoring described in the section 4.5.

Requirement Number	The knowledge model shall	Priority
UPD-1	allow for the update of all details that are stored within its instances in line with changes to legislation, good practice guidelines and standards	M
UPD -2	maintain a record of all updates as specified in the audit requirements	M
UPD -3	allow for the update of the model itself, reflecting changes to users and maintaining existing information stored in previous versions of the model	S
UPD-4	permit collaborative working for the authoring, editing and development of the model itself where appropriate, including editing, comments and publication	S
UPD-5	set a status of the Secutype model to indicate whether it is under development, published for use or no longer in use	S
UPD-6	maintain a version history of each Secutype as they are edited and updated	M

Table 8: Requirements for Specification of Knowledge Model

The Secutypes are intended to provide a series of blueprints from which policy and guidance for handling EHRs in research can be developed and disseminated amongst people involved in a given context of use. The next section provides the requirements for the policy tool use that this thesis proposes.

6.3.4. Secutypes Used for Working Practice

The Secutype model is used as a basis for developing an information management system that this thesis proposes should act as an information governance policy editor for establishing the details usually found in a policy and other governance documentation such as data sharing agreements for example. The requirements for this tool are provided in Table 9. POLED-1 represents a core element of the research work under evaluation and also has a basis in examples of EHR driven systems that are underpinned by an Archetype and Information Model described

in Chapter 4 and section 5.5 as well as the management of documentation described in section 7.5 of ISO 27001:2013. The source for POLED-2 includes ISO 27001:2013 section 7.5.2 and 7.5.3, and NHS guidelines such as Information Security Management and the Confidentiality Code of Practice, which require that the policies themselves can be updated to cover any changes in legislation, data sharing policies or working practice within a use context. A version history of the policy items would also be required so that a complete trail of which policy item versions were in use at a given point would be available.

Requirement Number	***The policy editor shall***	***Priority***
POLED-1	store and present all elements of an information security policy, data sharing agreement or contract according to the model specified by the Secutype	M
POLED-2	allow the update of details stored within instances of the Secutype model within a given context of use	M
POLED-3	log tool use in accordance with audit requirements	M
POLED-4	permit access to policies from authorised users only	M
POLED-5	allow the export of computable artefacts to assist with the configuration of security software tools such as access control and privilege management	M

Table 9: Secutype Based Policy Editing Tool Requirements

POLED-3 is related to GEN-11 in Table 2, where the sources for this requirement include the ISO 27001:2013 specification for ensuring the competence of members of the organisation, Confidentiality Code of Practice under the principle of "no surprises" for patients and the case of P against Finland at the European Court of Human Rights is pertinent here (HealthImaging, 2008), as P won the case primarily because the defendant could not show that her records were being protected in the way that they claimed. This is a strong indication that older versions of policy documents should be available and that an audit trail of policy preparation and access by users is needed. POLED-4 has a basis in ISO 27001:2013 under section 7.5.3 point b), which requires adequate protection for documentation.

The literature reviews and case studies also demonstrated that there was significant concern on the part of data subjects and the research community that

working practice was occurring in line with the expectations and requirements established by good information governance. Conversely, successful prosecutions by the ICO have relied on effective audit of interaction with patient records. This provided the basis for developing requirements around auditing any interactions with the information governance details on the part of authors, including creation, amendment and access to review. These requirements are included in Table 10.

Requirement Number	***The knowledge model shall***	***Priority***
AUD-1	capture details of interaction with any information held within instances of the knowledge model, including when items are added, amended, viewed and updated	M
AUD-2	store the time any items held within the instances of the knowledge model have had any interactions, including calendar date and time	M
AUD-3	capture details of who interacted with these elements of the policy items	M
AUD-4	capture details of the purposes for the interaction with the policy tool items	M

Table 10: Assurance and Audit Requirements

The consideration of these more refined conclusions and an analysis of the requirements listed in Tables 2 - 10 have resulted in the development of use cases and sequence specifications in the following sections. The Secutype model needs to be implemented and instantiated for use within an appropriate information management system to meet the requirements specified above. The results of the research work therefore also include a set of use case specifications to handle the use of the Secutype model to create, update and review policy items for advisory in handling sensitive information, which can be found after the knowledge model class diagrams. They have also informed the creation of a set of data templates that represent the kinds of information would need to be captured to inform domain knowledge so that the required guidance can be developed and achieved. These templates have been used for further analysis and to develop the knowledge model itself, represented as a set of Class Diagrams.

6.4. Secutype Driven Use Cases

The design and development of the knowledge management framework has included an analysis of use cases that involve instances of the knowledge model to allow for narrative based policy and good working practice guidance to be created, managed and used to offer advisory on how to handle EHR information when shared for healthcare research. This section provides the use cases that have been developed to represent the planned use of the Secutype and guide its implementation. Section 6.4.1 provides the use case for specifying and managing the Secutype model. Section 6.4.2 specifies the use case for using the Secutypes in practice through a the Secutype model driven policy editing tool that offers advisory on handling EHRs in clinical research. The actors identified in the research to date for managing and developing Secutypes and using an information management system to specify policy according to the Secutype constraints would include anyone involved in the processing and management of information. It is clear from the research work that whilst senior members of an organisation should own the risks involved with processing sensitive information or hold the role of a Caldicott Guardian, good governance practice requires that as many stakeholders as possible be involved in the development of policies. The actors in these use cases could therefore represent any person who is part of a research organisation or project.

6.4.1. Secutype Authoring, Editing and Management

It is clear from the research work to date that it must be possible to create, edit and update the Secutype models in order to allow for the creation of a collection of records pertaining to information governance requirements within a specific context of use. Research into the EHR work in this area has uncovered a series of editors and scripting languages that are used to achieve this purpose, rather than manually writing static high-level language implementations of specified models. This has allowed examples of collaborative working to specify a commonly agreed blueprint between clinical domain experts for how EHRs should be stored, one which can be grown into a library of domain model implementations. Given the requirements for developing a similar blueprint that can be updated and adapted

for information governance details and the collaborative engagement that is needed, the requirements for a similar facility to specifying the Secutypes is apparent. Figure 27 provides a use case for this Secutype editing example.

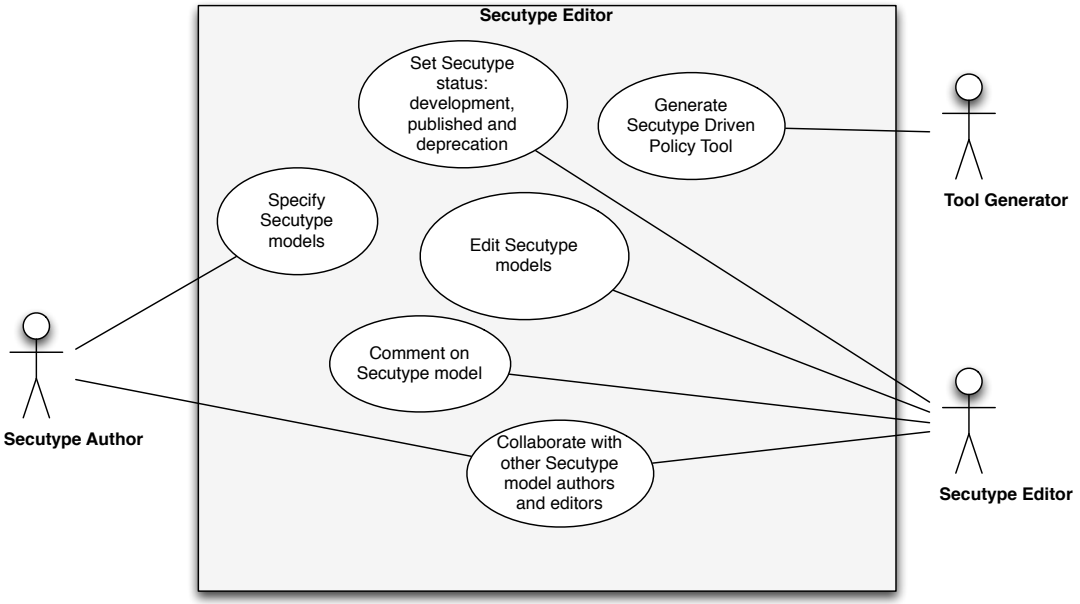


Figure 27: Secutype Specification, Editing and Management Use Case

Three actors are presented in this use case: a Secutype Author, who would specify a new Secutype, and a Secutype Editor, who would update that Secutype as required. These Actors could be the same or different people. They would collaborate with one another to develop the Secutype models and specify them. They would also be able to apply comments to the different versions of the Secutype as they are edited and evolved. Once the Secutype had reached an appropriate level of specification, it could be used to guide the development of an information management tool that would encapsulate information governance policies, guidelines and good working practice.

6.4.2. Use of Secutypes in Practice: Knowledge Model Driven Policy Editing Tool

The policy-editing tool envisioned by the research work would be generated from the Secutypes authored as per the Use Case described in Figure 27. This would provide the tool for use in practice, whereby policy (or any other information governance control specifications) could be entered on to the information system

and thereafter updated, used to guide behaviour and assert the specified controls. Figure 28 illustrates this use case.

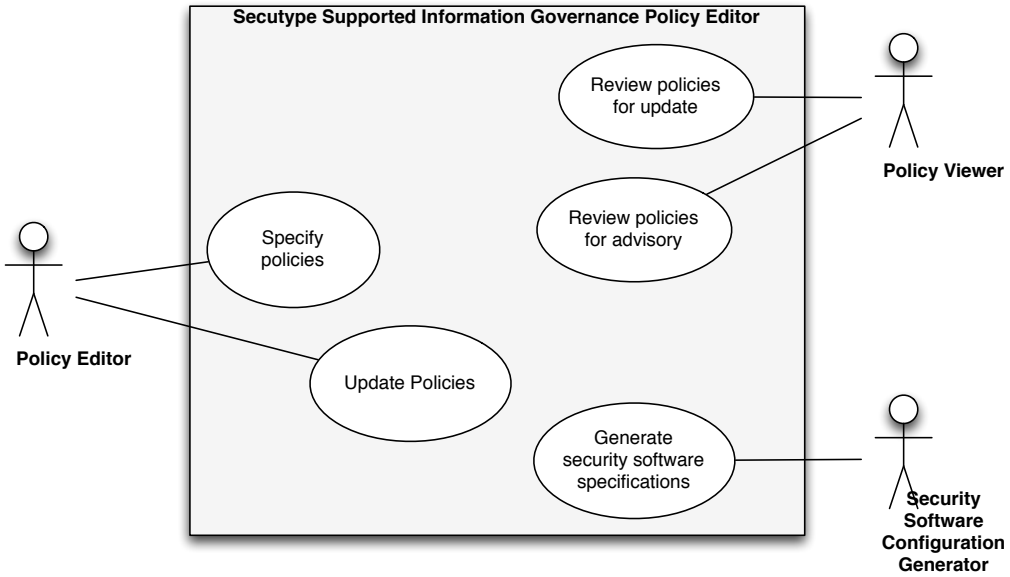


Figure 28: Policy Tool Use Case

Figure 28 shows three actors, who could be any individual working with EHR data in a care or research context of use as with the Secutype generation use case. Policy Editors such as senior information risk owners, Caldicott Guardians, principal investigators or researchers both specify and update their policies according to their information security management system, including analyses of risk and safeguards for a given project or other endeavour. The policy viewers who could be the same individuals or any person or system whose activities must be guided by the details in the policy, data sharing agreements or code of practice.

6.5. Knowledge Management Sequence Diagrams

The steps involved in developing Secutypes should be easy to understand and accomplish. This section describes the proposed steps using high level sequence diagrams to illustrate the development of the Secutypes through to their use in deploying a policy editing and review tool. Figure 29 shows this diagram, which illustrates how the Secutype would be developed using an editor not unlike those introduced in section 4.4.3 for Archetype development. Once specified, the Secutype could be exported using a scripting language as Archetypes are using ADL or XML. These artefacts would subsequently be used to guide the

development the policy-editing tool according to the Secutype constraints, which could then be used by the actors for advisory and system configuration as shown.

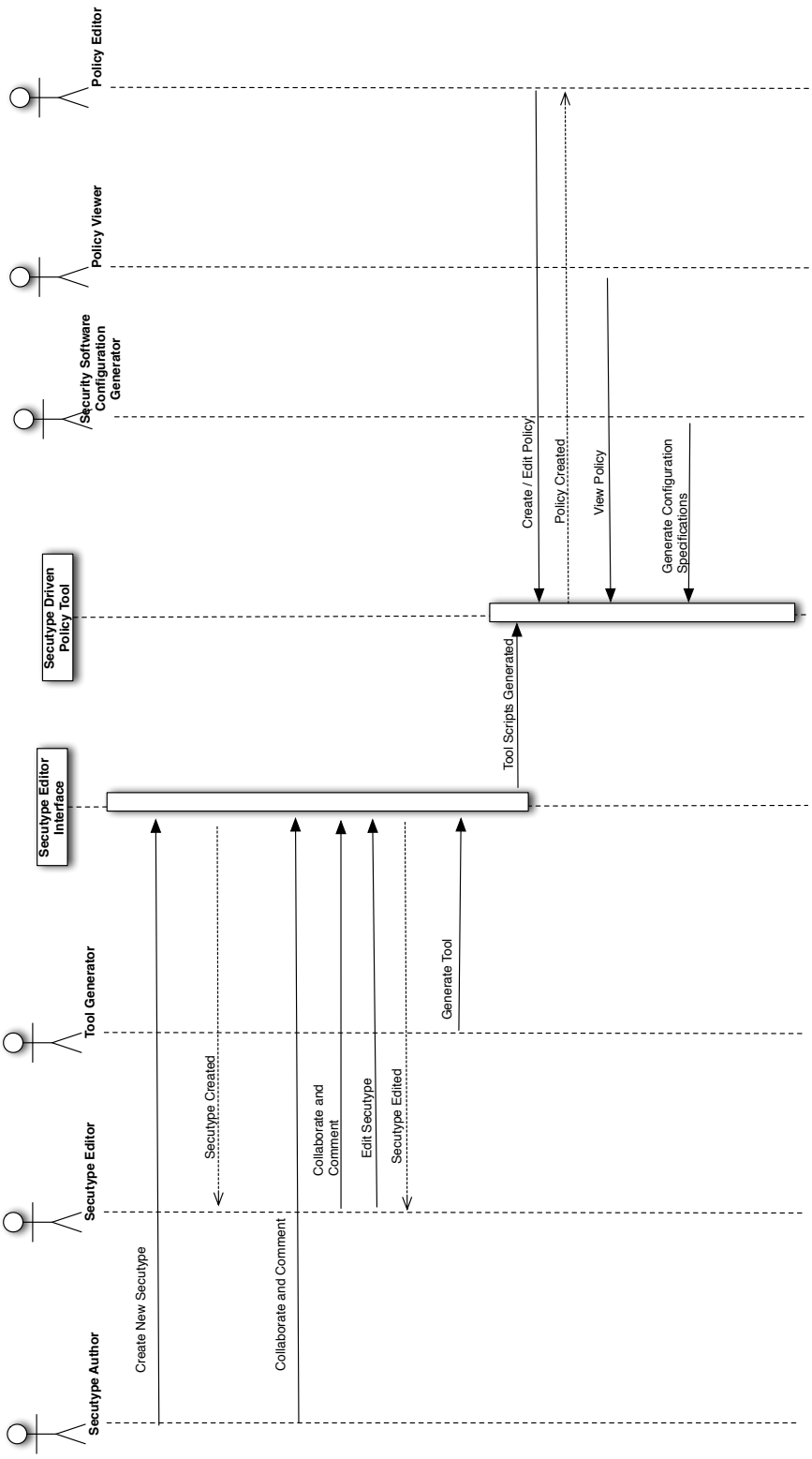


Figure 29: Sequence Diagram for Secutype Specification and Policy Tool Implementation

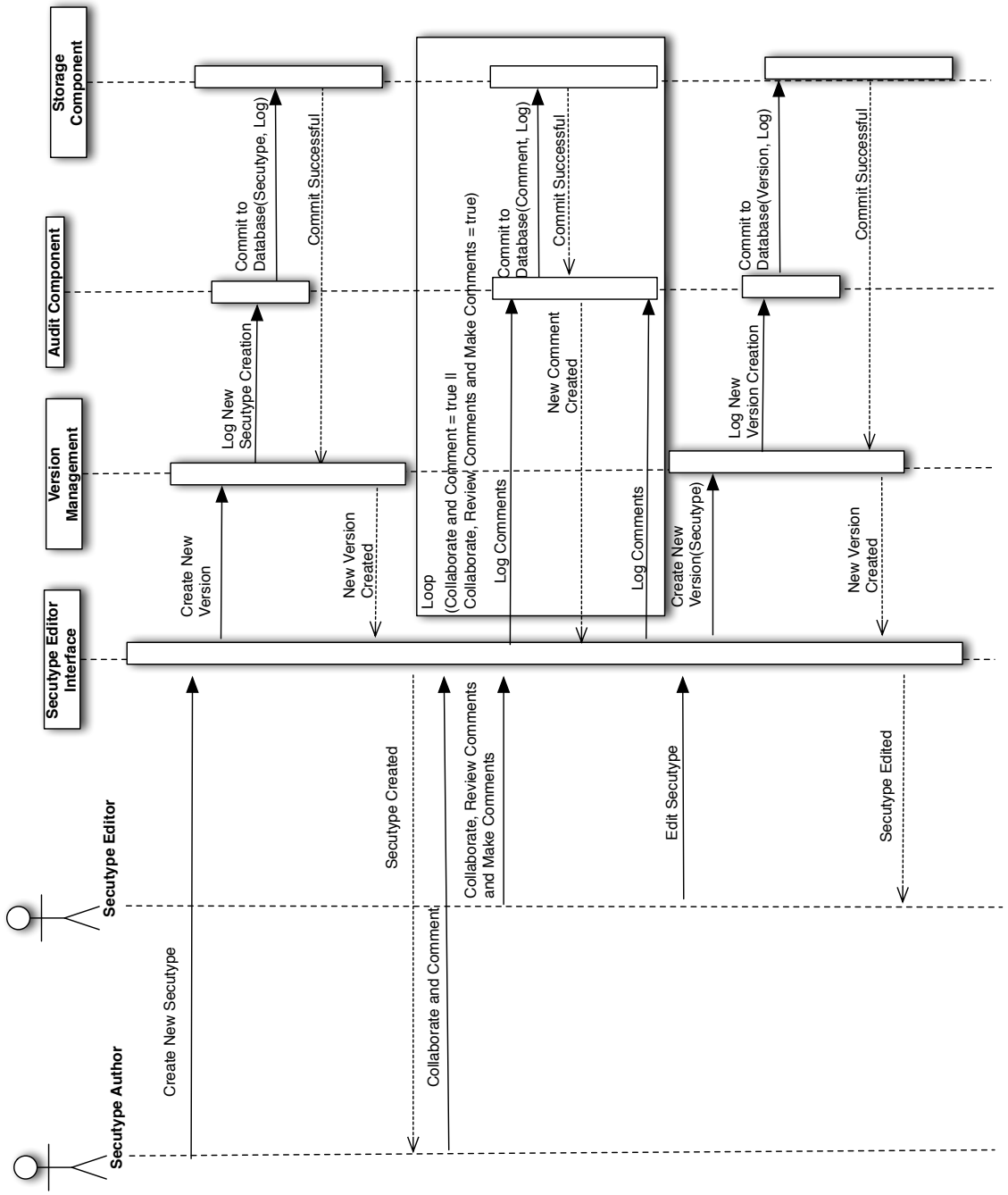


Figure 30: Lower Level Sequence Diagram for Secutype Specification

Figure 30 provides a lower level, more detailed sequence diagram of the proposed process for developing Secutypes. Here the Secutype Author creates a Secutype, and collaborates with other Secutype editors to evolve that Secutype to a commonly agreed specification, using a comments system that is tied to each version of the Secutype as it is developed. This step is iterative and would be repeated as shown in the diagram loop and the versioning component would allow

for a complete record of the evolution of a particular Secutype as it is refined and developed. Once agreed, the Secutypes could be exported into an appropriate format to guide the development of the policy-editing tool. Once the policy-editing tool has been developed, its proposed use is specified in the sequence diagram represented in Figure 31.

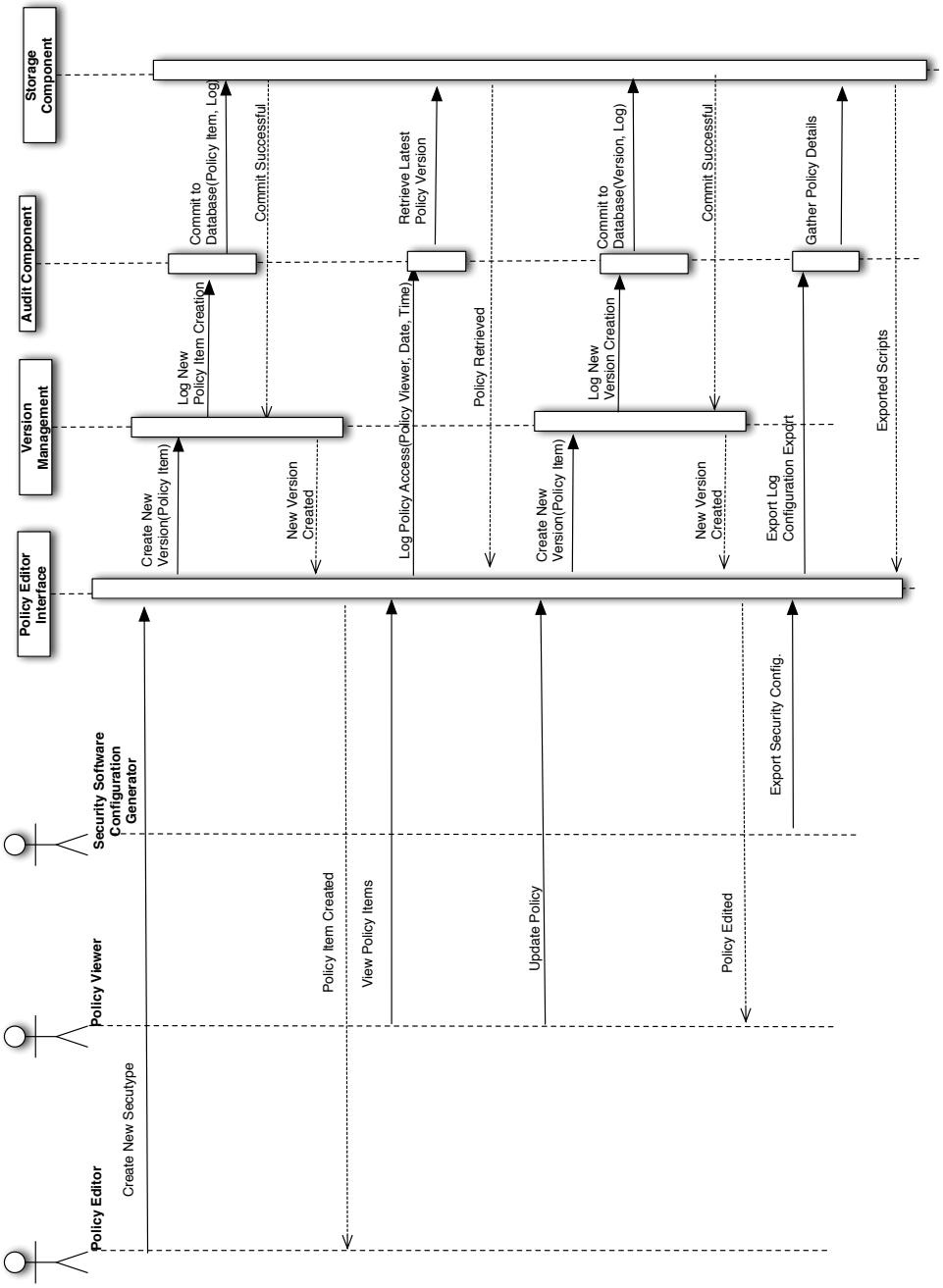


Figure 31: Lower Level Sequence Diagram for Policy Editing Tool Usage

The proposed use of the policy editing tool centres around the specification of policy items according to the information structure developed using the Secutype. Once edits have been made to the policy, viewers of the policy will be able to take a look at it, as well as offer feedback for its update. The policy will be updated periodically to handle changes in information use. Policy Editors and Policy Viewers could be the same individual depending on their roles within a particular context of use, and would need to be able to update the policy according to required changes in working practice or any issues with the policy that may arise. A versioning component is also included so that the evolution of particular policy items or components therein can be recorded and reviewed where needed. A logging component would capture details of all the interactions with a particular policy, including editing, access for viewing and update, which would be used in an audit or potentially presentation in legal proceedings. The tool would also offer a feature to export any computable policy items to security software systems, for example and access control or privilege management system.

6.6. Templates for Secutype Model Requirements Analysis

The requirements for the knowledge management framework have been described in the previous sections and illustrated by use cases and sequence diagrams. This section describes the process of designing the Secutype model itself. The requirements have been further analysed to provide a set of generic templates that express in more detail the information that needs to be captured and managed according the Secutype model. The primary goal of the template structure is to refine these needs so that a model for implementation can be designed and implemented.

The specification uses these templates and presents a generic UML model for the Secutype. Tables 11, 12, 13 and 14 specify the templates, which have been used to generate the Secutype class diagrams presented in section 6.7 to distil the required information into the constituent data items. The templates have been given a heading structure, the first being a context of data use; the next is how a request for data is received and the kinds of information that are needed in order to assess it; following the request is the template for a data release, which is very

similar to a request; the template for a control assertion has then been provided; and after that, a template of information to describe involved parties.

6.6.1. Template for Context

The template for the context of use is presented in this section. The context of use is broadly synonymous with a research or surveillance project, an endeavour for which a pool of clinical data is being held and requires access by people and processes to achieve its goals, the legal basis for pursuing these goals and what activities are authorised, as described in the requirements section Table 3. The template is described in Table 11.

Item	Value
Initiative Details	Describes the initiative, and what its purpose is
Location	List of locations where the initiative occurs and the contributing institutions / organisations
People involved	List of named individuals who are stakeholders within this context
Assets	Hardware, software, data items
Authorisation	The legal authorisation under which the initiative is being run
Authorised activities	List of activities that have been authorised in this context of use; this will include data release.

Table 11: The Context Template

6.6.2. Template for a Request:

The template for a data release request uses the following format. It is based upon the general pattern that occurs during research projects when new uses of the data are requested, the general format of applications for Section 251 exemption and

ethics committee approvals, as well as clinical care models for point of care data uses; it has used the Care Record Guarantee and the Confidentiality Code of Practice to specify the items and values:

Item	Value
Data items sought	List of the data items, or repositories that will be releasing data
Reason for use of data	List of reasons depending on the context of use and goals of an initiative
How data will be used	List of activities that will be performed with the data, with corresponding software and hardware use
Who will use the data	List of people who are part of a particular initiative
Where data will be used	List of locations of use as per context
How data will be retained as a result of this use	Details of whether data will be retained or not, and if so, where, how, and for how long if known

Table 12: The Data Request Template

6.6.3. 4.4.3 Template for Data Release

The data release template is a significant element in the knowledge management framework. It is here that the release can be specified, and specific controls are enacted based on the details that are specified

Item	Value
Data Resource	A list or single data item defined either as an information resource / repository, or specific EHR Archetype node / element item

Requesting Agent	Who, or what is requesting the data be released
Agent to which data will be released	Who, or what the data will be released to
Purpose for release	Why the data release is being requested
Legal Basis for Release	The legal basis for the release of the data
Data Release Method	How the data release is expected to occur

Table 13: The Data Release Template

6.6.4. Template for Controls:

The details of the control itself are captured in this template. The actual controls and values will be determined by the context of use and data release details. Table 14 presents the Control template:

Item	Value
Target Data Item	A list or single data item defined either as an information resource / repository, or specific EHR Archetype node / element item; this could include flat file resources that contain data (as a comma separated value or xml file, for example).
Control Type	List of available control types, currently data release control, audit control,
Control Specification	The exact control depending on type of control, target data or resource item, resources used and activity of use.

Scope	Whether the control is one that can be deployed and asserted in a live server setting, i.e. must be manually applied, or whether it applies to general guidelines for a whole resource.

Table 14: The Control Template

6.6.5. Template for People involved:

Table 9 presents the template for people involved in a given context. The details of role and access credentials will be specified here and will also determine what data is released and what controls can be applied:

Item	Value
Identity	Identification of a person involved in an initiative
Affiliation	Affiliation of person, to be supported by contract of employment and job description
Role	Job title, role within the initiative
Required Access	Access to Initiative Resources

Table 15: The People Involved Template

By analysing these templates with a view to making them computationally usable, they have been developed into the Secutype and are represented as Class diagrams using UML conventions in the next section.

6.7. Class Diagrams for Knowledge Models

The templates have provided a means to analyse the collected requirements and to propose a model for the Secutype that is designed to capture the information that is needed to assist information governance good practice. The analysis has helped

to design a proposed model whereby a context of use can be specified and allow for components of the Secutype package to be held within that context, representing the policies, data sharing agreements and any undocumented good working practice that are used to inform appropriate behaviour when handling EHRs or data derived from the records. Secutype models for data release controls have been prepared in order to evaluate the thesis. Other Secutypes for Audit, Risk Assessment, Access Controls and Data Integrity, are proposed as other types of Secutype to be explored in further work.

6.7.1. Secutype Model Package

The Secutype model is presented in Figure 32. It has been developed to capture details about a series of information governance related concepts identified in the requirements: safeguards, activities, information assets, associated protection mechanisms and the details about purposes for using the data.

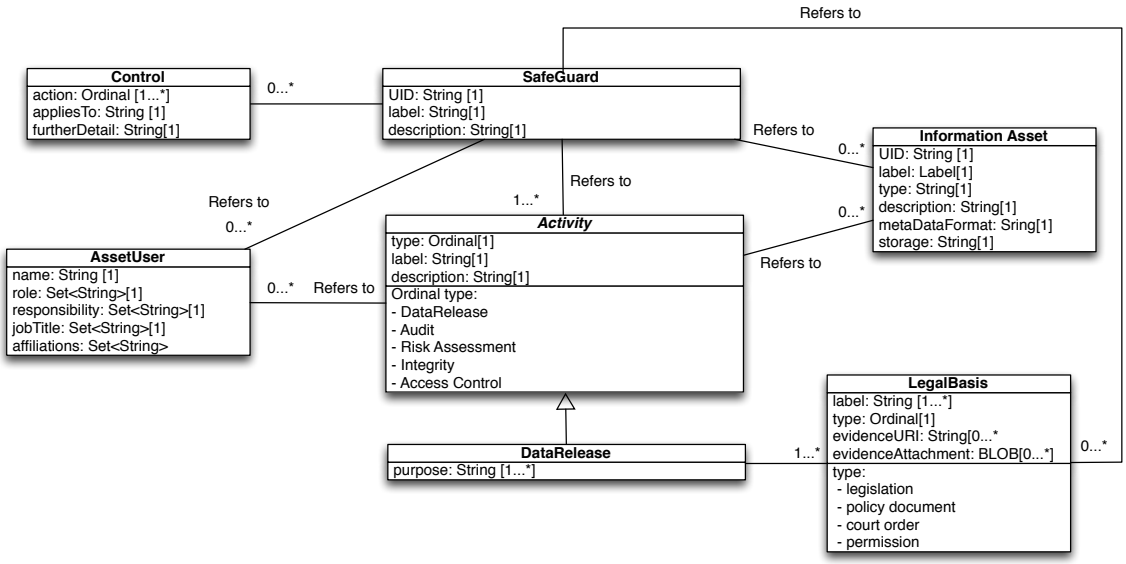


Figure 32: UML Representation of the Secutype Model

The Activity class is an abstract class, which is currently extended by the DataRelease class. The Activity class is intended to encapsulate a set of other activities that require security and governance management, including risk assessment to aid risk analysis, audit and integrity checking. These activities are proposed as further work and are discussed in Chapter 10. The Activity class provides a field to refer to Information Assets that are used for a particular

purpose for which Data Release is needed: the Information Asset is currently informed by the data resources that are held within a given Context, and contains details as represented by the requirements listed in Table 5.

The Activity class also has a reference to a series of Safeguards. Safeguards are proposed as a means of encapsulating details about the controls that are needed to protect the Information Assets as they are used by Asset Users in accordance with a series of Legal Bases. The Safeguard has been designed to meet the requirements listed in Table 7, where it includes a reference to the Activities, Asset Users, Information Assets and Legal Bases that may be what a particular Control or set of Controls refer to. The Safeguard has been designed to include these elements, which relate to what the controls refer to as well as the reasons behind them as a one stop reference for why they are in place and how they should be applied.

The Asset User class has been defined to represent the people involved in a particular use context and has been developed to meet the requirements listed in Table 4. Asset Users will be involved with Activities as well as have Safeguards applied to them both in a group and to particular Asset Users. It is possible to assign Asset Users to groups under the affiliations field. The Asset User class could be extended to include examples of software that processes information, though this is currently regarded as an Information Asset in this model.

The DataRelease class extends the Activity class and includes details on the purpose of a given Activity. Purposes are based upon the list provided in ISO TS 14265: 2012 standard for Health Informatics - Classification of purposes for processing personal health information (International Organisation for Standardization (ISO), 2011). Both DataRelease and Safeguard classes include a reference to a Legal Basis. This class has been developed to meet the requirements specified in Table 6, where it includes a reference to a type of Legal Basis according to a set of pre-set values, and a means to refer to one or more links to a documented set of evidence of a legal basis that may be available online, or to upload one or more files that represent this legal basis.

All the classes have a reference to a Label attribute, which could represent the UID attribute and has been designed based on observation and discussion with stakeholders across the various case studies about how best to provide a means to

identify a concept beyond its unique identifier or making provision for a simple name attribute. The Label is intended to provide a means to apply a set of natural language references and synonyms, or indeed to assist with translations for international use. The idea was established when considering concept codes in clinical practice, where multiple coding schemes exist (SNOMED, ICD etc.). Whilst there has been no identification of an equivalent for a coding scheme within the information security domain, the potential for this coupled with the inconsistency with how concepts may be described and interpreted has led to the use of the Label to manage this area. In addition, it is possible that the Label attribute will contain the UID for the given object, and this design is currently under a review.

6.7.2. Context Model

The Context model shown in Figure 33 has been designed to meet the requirements set out in Table 3. This represents the details that should be stored about any defined context of use within which information assets are being used to achieve a set of purposes. The *type* attribute provides a list of possible kinds of use contexts, including Research Project, Clinical Trial, Department or wider Organisation. The context can have details applied to it, including the contact details of the institution that it is part of, as well as a unique identifier (which could be a grant number or project identifier, for example).

The context of use was designed separately from the Secutype package because it represented the container that would hold the details that the Secutype model itself would represent. This has been designed to define the scope within which the policy items would be used and relevant. As part of future development, interaction between the different contexts of use where appropriate (for example, in the case of a collaboration between existing projects) will be explored.

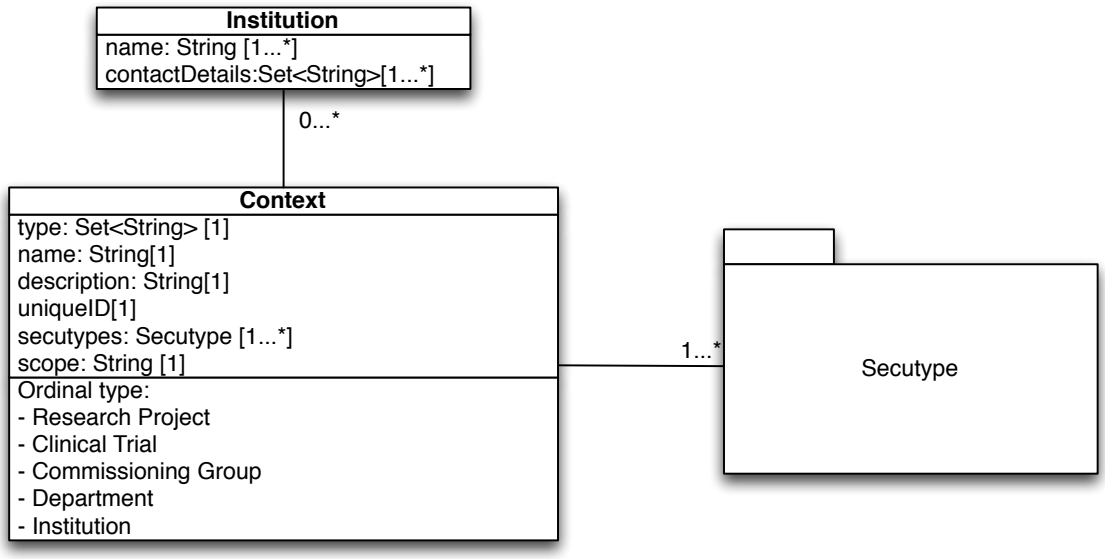


Figure 33: UML Representation of the Context Model

6.8. Summary of Requirements, Development and Design

This Chapter has provided a discussion of the approach taken to design and develop the knowledge management framework. The requirements for the knowledge management framework have included those for the knowledge model and a set for an editing tool for this knowledge model, the use of the model artefacts as part of an information governance policy-editing tool and the audit and logging requirements to provide some assurance that good practice has been adhered to.

The requirements have been supported by a series of use case analyses and analysed by developing a series of templates to represent the knowledge model components. Sequence diagrams represent the design and development of the editing tools and the knowledge model has been designed from this process. The next chapter describes the implementation of the knowledge management framework, guided by this requirements gathering, design and development work.

Chapter 7. Knowledge Management Framework Implementation

This chapter describes the implementation of the knowledge management framework designed and developed as described in Chapter 6. The Secutype and Use Context models in Figure 32 and Figure 33 have been designed to capture the information requirements for delivering a solution to manage information governance within a context of information use. Other requirements have been specified for authoring Secutype instances as well as using them within the solution to guide information governance in working practice. The development of these other components and the implementation of the Secutype itself were carried out in accordance with the observations made about EHR system development and deployment within the case studies. These case studies involved the author in the development of a new constraint model specification that was developed to simplify the process of editing Archetypes and deploying EHR systems.

This new constraint model was called the *Pattern* and designed in partnership with the EHR Team at the Centre of Health Informatics and Multiprofessional Education (CHIME) whilst working on the Cortext, EHR4CR, DeBugIT and Cardiovascular Application projects described in section 5.5. These constraint models would need to be applied to a set of EN 13606 reference models so that an EN 13606 compliant record server could be developed. The requirements for the Secutype as specified in section 6.3.2 and its editor as specified in section 6.3.3 were consistent with those for specifying the Pattern models for the next generation of EHR servers that were being developed for these projects. This represented an opportunity to align the implementation of the Secutypes with a reference model that had been ratified by ISO so that the requisite data could be captured and organised into a meaningful representation of information governance domain concepts to users.

7.1. Secutype Specification using the Pattern

The author decided to include the Secutype editing requirements into the development of the Pattern model, which would therefore enable it to represent the specification of the healthcare record domain concepts and the information governance domain concepts at the same time. The Pattern has been developed with the ability to specify the constraints specified by the Secutype. The class diagram of the Pattern is provided in Figure 34.

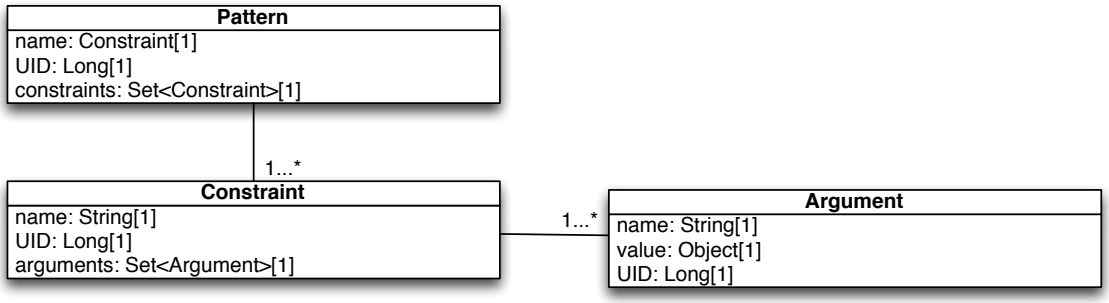


Figure 34: Pattern UML Model

The Pattern contains a field for its name and unique identifier, as well as a set of Constraints. The Constraint class is made up of a set of Arguments, with a name and UID. The Pattern encapsulates a series of Constraints, whose Arguments are given concrete Values to help represent clinical and information governance concepts according to each requirement. This was achieved using the dual factor modelling approach described in section 4.4, where the Pattern specifies the Constraints on an information / reference model implementation.

The EN 13606 Reference Model has been used to provide the information elements for the Patterns because its core information elements are the most generalisable, representing implementations for all basic data types within the most generalisable structure out of the EHR Standards. For example, the openEHR information model structure is categorised according to the recording of healthcare information whereas EN 13606 was produced as a means to communicate extracts of an EHR between communicating EHR systems. This lends a generalisable quality to the EN 13606 reference model, making it a candidate model to represent information items for information governance concepts by applying the record component structure to the Secutype model.

The EN ISO 13606 Reference Model has been used to structure the presentation of information governance concepts as represented in the Secutype Model. Table 1 lists the record component structure for EN ISO 13606, where the author has considered the applicability of each of the components to represent the Secutypes. The Pattern allows the specification of which of these record components must be used to Implement and represent the Secutype model. The record components provide a structure to the representation of domain concepts, where the information types are used to store data and the structure helps to contextualise that data in a meaningful representation. A full description of the record components used to implement the Secutype is provided in section 7.3 along with some examples of the Secutype Patterns to illustrate an implementation of this model.

7.2. Editing Secutype Patterns

Patterns needed to be created and managed in line with the requirements listed in Table 8. This led to the development of the Pattern / Secutype editing tool called *aruchi* (アルチ). This tool has been described by the author in a publication regarding the expression of security policy in EHR systems (Lea et al., 2009), which was presented at the World Academy of Science, Engineering and Technology held in Tokyo in 2009. This publication shows early examples of how the Secutype could be expressed through the *aruchi* tool, which featured a series of authorship details, publication statuses and versioning as specified in the requirements listed in Table 8.

aruchi was updated to include the ability to export the Secutype Patterns into a scripting language so that it could be used to assert the constraints on how an information management tool could be developed. The work on the Pattern as part of the case study projects saw an implementation of the Pattern design using the Scalable Language or *Scala* as the scripting language for the Secutypes. Scala is an object oriented and procedural language that has been developed to allow for succinct expression and inheritance based upon a series of procedural functions (Odersky et al., 2010). Scala was chosen due to the ease with which Patterns could be specified and generated by *aruchi*. The generated scripts could then be easily

worked into a framework for generating the structure and components of information management tools, as observed in the deployment case studies described in section 5.5. The next section describes the Secutype Patterns and provides some examples to illustrate the implementation of the Secutype and its expression using Scala scripts. This introduces how these scripts are used to develop information management systems and the choice of Scala and its use to help develop information management tools discussed in section 7.4.

7.3. The Secutype Patterns

The Secutype has been implemented using Scala scripts in accordance with the Pattern Constraint model, which allows the representation of the Secutype classes using the Reference Model specified in the EN ISO 13606 standard. The complete set of Scala scripts that the author has developed can be found in Appendix 5. The use of the EN 13606 record component hierarchy allows information to be organised according to a standardised hierarchical structure as described in Table 1, which in turn allows the Secutype concepts to be presented automatically in a screen for users. The author has used the Composition, Entry, Cluster and Element record components to specify the Secutypes.

The Composition has been used to convey a single meaningful contribution of information items to an information governance policy, much as it is used to convey a single, medico-legally meaningful contribution to a clinical record. The Composition sits higher up in the hierarchy of record components, and is a container for the Entry, so that the viewing, editing and storage of concepts recorded in a composition are represented in a way that is conformant with the specified constraint model and makes sense to users. The Compositions for the information governance concepts are Safeguards, Activities, Information Assets, Asset Users and Legal Bases.

The Entry has been used to define a single instance of the information items that form a record according to the specification of the Composition. It represents the information items that form the individual Safeguards, Activities, Information Assets, Asset Users and Legal Bases. The EN ISO 13606 standard has specified the Entry as well as a Section (which allows the addition of data under specific clinical

headings) to provide further abilities to categorise the information within a clinical context. During the iterative process of developing Secutypes according to the EN ISO 13606 record structure, the author has not found a need to add additional headings to the presentation of the information to meet the Secutype requirements, however the facility remains to add sections should the need arise.

The information items themselves are further grouped by Clusters and are contained within Elements. The Cluster is lower in the hierarchy of the record component and is a means to group together multiple Elements where it is useful to associate those Elements within a Composition. Elements themselves sit at the lowest level of the record component hierarchy and are designed to hold data items. The available data items from the EN ISO 13606 model are a candidate set described by Sun et al (Sun et al., 2012). This section describes how the Secutype model is represented by these record components, listing each of the implemented Secutypes according to the hierarchy of record components that is available. The Secutype and Control specifications offer examples of the Scala code to explain how that has been used to develop the Secutypes.

7.3.1. The Safeguard

The Safeguard class as shown in Figure 32 has been specified as a Composition using the Pattern model, which is made up of several Elements and a Cluster to represent a Control within the associated Entry. The Elements include a label attribute to easily identify the Safeguard in question and a description attribute, which allows for a longer specification of what the Safeguard has been specified to achieve. The Safeguard includes four reference fields, one for the legal bases upon which the Safeguard is based, if any, one for the activities that it has been specified to control, one for the Information Assets it refers to and one for the Asset Users for whom this Safeguard is intended. It is also possible to specify which types of asset a Safeguard may be applied to, a particular metadata format, be it relating to a particular EHR standard or item within a particular clinical Archetype, and a hardware storage attribute for which the Safeguard applies. The author has specified the Pattern to allow multiple values of these attributes, since the Safeguard may be applied to none or many of these. The Activity, Information

Asset, Legal Basis and Asset User references are linked to actual instances of these Activities that should have been added to the policy editing tool described in section 7.5. The following script provides an example of the Scala script used to specify the Safeguard Secutype:

```
//Generated by aruchi on Wed May 16 11:21:13 BST 2012 for nathan
def Safeguard : (Long, Pattern) = {
   var p = new EN13606Pattern(ma)
     .setDateLastVerified(new java.util.Date(1337087473670L))
     .setDateOfIncorporation(new java.util.Date(1337087473670L))
     .setDefinitionProvidedBy("Nathan Lea, CHIME, UCL, UK")
  .setDescription("en_GB", "Defines the Controls within a specified Safeguard as defined by an information Security Policy. ")
     .setEntry()
     .setLibraryPath("secutype")
     .setPatternIdentifier("Safeguard")
     .setPatternName("Safeguard")
     .setPublicationStatus(2)
     .setVersion(1)
     .addItem(labelID, "Safeguard Label", 1, 1)
     .addItem(descriptionID, "Description", 0, 1)
     .addItem(legalBasisReferenceID, "Legal Basis Reference", 0, null)
     .addItem(Control._1, "Control", 0, null)
     .addItem(ActivityReference._1, "Activity Reference", 0, null)
     .addItem(informationAssetID, "Information Asset Reference", 0, null)
     .addItem(assetTypeID, "Asset Type", 0, null)
     .addItem(hardwareStorageID, "Hardware Storage", 0, null)
     .addItem(assetMetadataID, "Metadata Format", 0, null)
     .addItem(assetUserID, "Asset User Reference", 0, null)
     .setViewOrder(
        Array[String]("Safeguard Label", "Description", "Legal Basis Reference", "Activity Reference", "Information Asset Reference", "Asset Type", "Hardware Storage", "Metadata Format", "Asset User Reference", "Control")
      )
      .setPreviewOrder(
        Array[String]("Safeguard Label", "Description")
       )
     .getPattern();
```

```
    p.publishable = true;
    p.username = "nathan";
    return (lastID + 1, p);
  }
```

Code Excerpt 1: Safeguard Secutype Scala Coded Pattern

The Scala program excerpt provided in Code Excerpt 1 represents the definition of the Safeguard Entry. It is defined as in a class structure with a Pattern name and unique identifier as shown in the Pattern class diagram in Figure 34 to store it within the library of Secutypes that have been developed. The script instantiates a variable as a new EN13606Pattern (p), where the EN13606 class specifies the constraints on the structure of the pattern, for instance that a Composition can hold an Entry, which itself can hold a Cluster and Element, but that an Element could not hold a Composition. This also includes the specification of common attributes to all Patterns, including authorship details using the DefinitionProvidedBy attribute, a description, as well as dates when the Pattern was last verified and incorporated into the Secutype library during the collaborative review process described in Figure 30. A publication status has also been provided in line with the EN ISO 13606 statuses, as well as the specification as to whether the Secutype Pattern is publishable or not. The identifier is also automatically added by taking the last identifier used in the library and increments the identifier by one. The version of the Secutype specification can also be set using the setVersion attribute, which is automatically set using *aruchi*. These features satisfy the requirements specified in Table 8 for editing and management of the Secutypes.

Where arguments contain ID appended, as illustrated in Code Excerpt 2, this represents a String identifier for a Pattern that has already been added to the library. This code excerpt has been highlighted in Code Excerpt 1.

```
.addItem(labelID, "Safeguard Label", 1, 1)
.addItem(descriptionID, "Description", 0, 1)
```

Code Excerpt 2: Reference to Previously Added Pattern Identifier

The setEntry() method call specifies the Safeguard Secutype as an Entry, which allows each of the items to be added to that Entry. For example, the Control can be added as per the following method call, as highlighted in Code Excerpt 1:

```
.addItem(Control._1, "Control", 0, null)
```

Code Excerpt 3: Addition of Control Cluster to Secutype Entry

where the arguments to the addItem method consist of a Pattern Identifier (in this case Control._1, a common name for that Item ("Control") and a multiplicity, where '0' states that the minimum number of Controls is not mandatory, and 'null' specifies that there is no limit on how many Controls can be added, making the Control multiplicity zero or many within a particular Safeguard. This example illustrates one the features of the Scala programming language, which allows for tuples as return types. The Item class Control is also an example of a Pattern and defined in the same way as the Safeguard returning a (Long, Pattern) pair, the first part of which (._1) is supplied as the first argument to addItem. The remaining arguments are a String common Name (Control in this case) a minimum int (here '0') and a maximum int (here null and therefore applying no upper limit). The next subsection describes the Control Secutype and illustrates how its use within the Safeguard has been coded in Scala.

7.3.2. The Control

The Control class has been specified according to the Secutype package specification provided in Figure 32. It provides for a specification of an Action (which must be taken to apply the control), which AppliesTo a particular algorithm or specified for human behaviour, and Further Detail is provided should any extra arguments need to be applied. The Action is formed of a series of defined values, which can be updated and amended according to the context within which the action must be applied. For the purposes of this work, the author has defined three actions for security controls: Apply, Forbid and Permit. The author has proposed that these three actions are sufficient for the purposes of this work but this does not assert that the list is exhaustive. The values for this attribute have

been fixed to aid with the simplicity of specifying policy and to encourage a consistent expression for specifying controls.

AppliesTo has been defined to specify what the Action applies to. This also contains a list of predefined values, Access, Anonymisation Function, Behaviour, Blur Function, Release and Sensitivity Level, which the Action should be applied to. The FurtherDetail Element has been provided so that any additional details can be applied to a Control. An example of a Control specification using these terms is provided at the end of this section. The Control has been defined as a Cluster so that it can incorporate these three Elements to represent these data items. It has been added to the Safeguard Entry and its Scala code is provided below to illustrate how it has been implemented:

```
//Generated by aruchi on Wed May 16 11:21:13 BST 2012 for nathan
def Control : (Long, Pattern) = {
  var p = new EN13606Pattern(ma)
        .setDateLastVerified(new java.util.Date(1337087473670L))
        .setDateOfIncorporation(new java.util.Date(1337087473670L))
        .setDefinitionProvidedBy("Nathan Lea, CHIME, UCL, UK")
     .setDescription("en_GB", "Pattern that defines a Control, which forms part of a particular Safeguard.")
        .setCluster()
        .setLibraryPath("secutype")
        .setPatternIdentifier("Control")
        .setPatternName("Control")
        .setPublicationStatus(2)
        .setVersion(1)
        .addItem(Action._1, "Action", 1, 1)
        .addItem(AppliesTo._1, "Applies To", 0, 1)
        .addItem(FurtherDetail._1, "Further Detail", 0, 1)
    .getPattern();
  p.publishable = true;
  p.username = "nathan";
  return (lastID + 5, p);
}
```

Code Excerpt 4: Control Secutype Scala Coded Pattern

The same attributes regarding the development, authoring and type of record component are available for all Secutypes, as in Code Excerpt 4. The Control has been specified as a Cluster as defined in the EN ISO 13606 standard using the .setCluster method., allowing a set of Elements to be added to that cluster so that it can be kept in this collection within the Composition. In this case, the three Element record components have been added to the Control, the Action, AppliesTo and FurtherDetail. This represents the three information items that the Secutype models define as what is needed to specify an information governance control. Code Excerpt 5 gives an example of an Element specified using the Pattern library, using FurtherDetail to illustrate.

```
 def FurtherDetail : (Long, Pattern) = {
 var p = new EN13606Pattern(ma)
       .setDateLastVerified(new java.util.Date(1337007349315L))
       .setDateOfIncorporation(new java.util.Date(1337007349315L))
       .setDefinitionProvidedBy("Nathan Lea, CHIME, UCL, UK")
    .setDescription("en_GB", "Any other details required to successfully apply this
control")
       .setElement()
       .setLibraryPath("secutype")
       .setPatternIdentifier("FurtherDetail")
       .setPatternName("Further Detail")
       .setPublicationStatus(2)
       .setStringWithLanguage()
       .setMaxLength(255, true)
       .setVersion(1)
   .getPattern();
 p.publishable = true;
 p.username = "nathan";
 return (lastID + 1, p);
}
```

Code Excerpt 5: FurtherDetail Element Scala Coded Pattern

The Scala code uses the method setElement to specify that this Pattern is intended to be an Element item in accordance with the EN ISO 13606 specification. This means that this record component can hold actual data values. The setString

method specifies that this Element would hold string values, and the setMaxLength attribute specifies that this string would have a maximum length of 256 characters. This Element, along with the Action and AppliesTo Elements, form the Control Cluster as specified in Code Excerpt 1.

7.3.3. The Activity

Activities have been defined as Compositions as they represent a core component for specifying information governance policies by understanding how to develop risk mitigation strategies. Activities may contain zero or more Information Assets and involve zero or more Asset Users. The Scala implementation provides the ability to refer to these Asset Users and Information Assets provided the details are available. In addition to these, the Activity can have zero or more specific purposes, listed as Auditing, Clinical Care, Clinical Trial, Commissioning, Education, Marketing, Medical Research or Public Health Surveillance. Code Excerpt 6 provides the Pattern used to define the PurposeType within the Activity Secutype. It shows that the Purpose is defined as a Named Property: this allows for the eight types of purpose to be specified elsewhere within an editing application and used as a preset list of values when Activities are being defined. This list was derived from ISO TS 14265 (International Organisation for Standardization (ISO), 2011) and can be updated and edited, and is the same type that has been used to define the Action and AppliesTo fields in the Control specification.

```
def PurposeType : (Long, Pattern) = {
   var p = new EN13606Pattern(ma)
      .setConceptDescriptor()
      .setDateLastVerified(new java.util.Date(1337088127404L))
      .setDateOfIncorporation(new java.util.Date(1337088127404L))
      .setDefinitionProvidedBy("Nathan Lea, CHIME, UCL, UK")
    .setDescription("en_GB", "A set of purposes as defined by a draft ISO Standard on Purposes for Healthcare Information Use.")
      .setElement()
      .setLibraryPath("secutype")
      .setPatternIdentifier("PurposeType")
      .setPatternName("Purpose Type")
      .setPublicationStatus(2)
```

```
      .setVersion(1)
      .set("NamedProperty", Array[Object]())
      .set("PropertyValues", Array[Object](
        "PURPOSE".asInstanceOf[Object]
      ))
  .getPattern();
  p.publishable = true;
  p.username = "nathan";
  return (lastID + 1, p);
}
```

Code Excerpt 6: Activity Purpose Scala Coded Pattern

7.3.4. The Information Asset, Asset User and Legal Basis

The three other core Secutype classes that have been implemented according to the Pattern model and are specified as single meaningful contributions to an information governance policy are the Information Asset, Asset User and Legal Basis. The Information Asset has been specified as per the class diagram specification in Figure 32, as has the Asset User. The Legal Basis has also been implemented, with the specification of evidence using a URI and Multimedia to provide a link to the legal basis documentation or the documentation itself as an uploaded document. In addition to this, the Legal Basis type has been implemented using a coded ordinal, where the list of possible values is specified in the Pattern itself. This approach is used when the list of possible values would be in line with an update to the core Secutype or Pattern as opposed to items that are likely to be context specific (like a particular de-identification algorithm) where a Named Property would be more appropriate.

7.3.5. Illustrative Example of a Secutype

To illustrate how the Secutype is expected to store information about information governance policy, the example provided by part four of the EN ISO 13606 Privilege Management and Access Control (British Standards Institute, 2009) is used as an example of how the Secutype Components would be defined. Figure 35 shows an excerpt from this standard regarding the management of privileges and access control to various elements of a patient's healthcare record along with a

description of the scenario. The background to the scenario and illustrations of the EN ISO 13606 model can be found in Annex A of the fourth part of the standard.

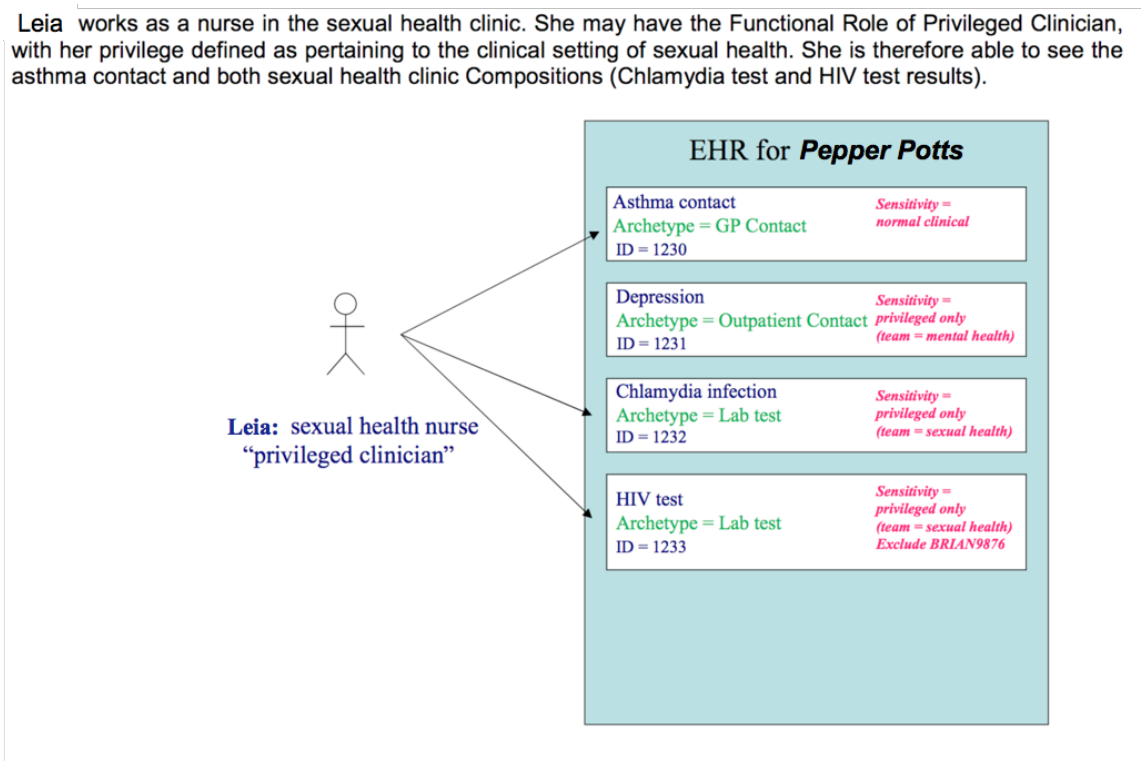


Figure 35: Example to Illustrate Secutype Specification from EN ISO 13606 Part 4: Privilege Management and Access Control

The specification of this scenario using the Secutype could be provided as illustrated in Table 16 to Table 25 and described below.

Secutype	***Details***	***Arguments***
Asset User	Full Name	Princess Leia Organa
	Job Title	Sexual Health Nurse
	Role	Privileged Clinician

Table 16: Asset User Example from EN ISO 13606-4

An Asset User Princess Leia Organa would be defined as in Table 16, given the job title sexual health nurse and role privileged clinician. Each component of Pepper Potts' EHR would be defined as an Information Asset: here there would be four, the

first, which is provided in Table 17, would be labelled as an Asthma Contact, with a UID of 1230, a description of "Archetype GP Contact" and a Metadata Format specific to the EHR standard that defines it. The hardware storage could be specified as unknown and the asset type would be EHR Component: Composition.

Secutype	*Details*	*Arguments*
Information Asset	Label	Asthma Contact
	Description	Archetype GP Contact
	UID	1230
	Metadata Format	EN ISO 13606 Extract
	Hardware Storage	Unknown
	Asset Type	EHR Component: Composition

Table 17: First Information Asset Example from EN ISO 13606-4

The Second, as shown in Table 18, would be labelled Depression with a UID of 1231 and a Description of "Archetype Outpatient Contact," with the Metadata Format, Hardware Storage and Asset Type as with the first example.

Secutype	*Details*	*Arguments*
Information Asset	Label	Depression
	Description	Archetype GP Contact
	UID	1231
	Metadata Format	EN ISO 13606 Extract
	Hardware Storage	Unknown
	Asset Type	EHR Component: Composition

Table 18: Second Information Asset Example from EN ISO 13606-4

The third and fourth Information Assets provided in Table 19 and Table 20 would be labelled as Chlamydia Infection with a UID of 1232 and HIV Test with a UID of

1233 respectively, with the other details the same as the first and second Information Assets.

Secutype	*Details*	*Arguments*
Information Asset	Label	Chlamydia Test
	Description	Archetype: Lab Test
	UID	1232
	Metadata Format	EN ISO 13606 Extract
	Hardware Storage	Unknown
	Asset Type	EHR Component: Composition

Table 19: Third Information Asset Example from EN ISO 13606-4

Secutype	*Details*	*Arguments*
Information Asset	Label	HIV Test
	Description	Archetype: Lab Test
	UID	1233
	Metadata Format	EN ISO 13606 Extract
	Hardware Storage	Unknown
	Asset Type	EHR Component: Composition

Table 20: Fourth Information Asset Example from EN ISO 13606-4

An Activity of care provision would be specified, as shown in Table 21, with a care purpose and a Legal Basis of consent that contained a digital version of the consent form signed by Pepper Potts for treatment. There could also be another Legal Basis linked to this Activity that included the consent form for the HIV test, as shown in Table 22 and Table 23.

Secutype	*Details*	*Arguments*

Activity	Label	Sexual Health Clinic Nursing Care
	Description	Nursing care provided in a sexual health clinic
	Purpose	Clinical Care
	Asset User Reference	Organa, Princess Leia
	Information Asset References	Asthma Contact (1230), Chlamydia Test (1232), HIV Test (1233)
	Legal Basis References	Consent (Care), Consent (HIV Test)

Table 21: Activity Example from EN ISO 13606-4

Secutype	***Details***	***Arguments***
Legal Basis	Label	Consent for Care
	Description	Consent for Care from Pepper Potts
	Type	Consent
	Evidence Attachment	Consent Form for Care in Sexual Health Clinic

Table 22: First Legal Basis Example from EN ISO 13606-4

Secutype	***Details***	***Arguments***
Legal Basis	Label	Consent for HIV Test
	Description	Consent for HIV Test from Pepper Potts
	Type	Consent
	Evidence Attachment	Consent Form for HIV Test at Sexual Health Clinic

Table 23: Second Legal Basis Example from EN ISO 13606-4

Several Safeguards are possible for this example. There could be one that granted access to Leia for the three record components that she had access to, which is provided in Table 24. This example would be accordingly Labelled and Described. It would add the three record components Asthma Contact, Chlamydia Test and HIV Test, and Activity of Care provision, as well as the Consent for treatment Legal Basis. It would then add a single Control with an Action of Permit which would use the Applied To argument of Access and specify Pepper Potts with a unique identifier

like NHS number in the Further Detail field to identify that it was Pepper Potts' record that this control applied to in order to specify Leia's access rights.

Secutype	***Details***	***Arguments***
Safeguard	Label	Access to Pepper Potts' EHR for Leia Organa
	Description	This Safeguard specifies the access to record components that Leia Organa may access for the purposes of Clinical Care
	Information Asset References	Asthma Record (1230), Chlamydia Test (1232) HIV Test (1233)
	Activity Reference	Sexual Health Clinic Nursing Care
	Legal Basis References	Consent for Care, Consent for HIV Test
Control	Action	Allow
	Applies To	Access
	Further Detail	Pepper Potts' EHR (NHS Number 123 456 789)

Table 24: First Safeguard Example from EN ISO 13606-4

Secutype	***Details***	***Arguments***
Safeguard	Label	Restriction of Access to Pepper Potts' EHR for Leia Organa
	Description	This Safeguard specifies the access restrictions to record components that Leia Organa may access for the purposes of Clinical Care
	Information Asset References	Depression Record (1231)
	Activity Reference	Sexual Health Clinic Nursing Care
	Legal Basis References	Consent for Care, Consent for HIV Test
Control	Action	Forbid
	Applies To	Access
	Further Detail	Pepper Potts' EHR (NHS Number 123 456 789)

Table 25: Second Safeguard Example from EN ISO 13606-4

An additional Safeguard that specified Leia as the Asset User and the Activity of Clinical Care as in the first Safeguard, and in this case the Information Asset of Depression would have a Control Added that used the Action of Forbid and the

Applies To argument of Access with the Further Detail of Pepper Potts' EHR (as in the previous example) to forbid Leia's access to the Depression record. This is shown in Table 25. Other Safeguards could be specified to allocate a sensitivity level to the record components and to assign Leia and her role access to any particular sensitivity level as well as particularly sensitive records. This would be an additional and possibly superfluous policy specification because the Secutype already provides a means to allow the access and enforce restrictions needed for particular EHR components pertaining to an individual for individual record components, rather than specifying general rules for allowing a group of role holders access to a set of record components. The access control mechanisms within EHR servers may need to be updated to support this level of specification. This is discussed under further work in Chapter 10.

These examples have illustrated how the Secutype model could be used to store details of the policy items for use by other systems to configure security policy, or to inform people on how to behave with sensitive information according to a specified policy or approved working practice. The next section describes implementation of the policy-editing tool to allow users to add, edit and review policy items constrained by the Secutype model.

7.4. Secutype Patterns Use to Guide Development of Policy Editing Tool

The literature review and case studies presented many examples of knowledge modelling and implementation of systems for managing healthcare records. These examples helped provide an approach upon which the modelling and development of information governance related records could be based, allowing for the Secutype models to be processed in the same way. By considering the constraint based models provided by Archetypes, the possibility of using the Pattern constraint model that could handle the clinical modelling requirements as well as those related to information governance and the Secutypes presented itself. The Pattern software package was developed to capture constraint model specifications and to translate those to inform the specification and generation of database schemata, user screens and business logic that provides screen

functionality and persists the data as required so that it can be stored in a record and retrieved. This would allow the Secutype model to be represented in a tool consistently with the requirements specified in Table 9. This section describes the process and the software that have been developed and used to specify and implement a tool to express the policy editor based upon the Secutype Knowledge Model.

The specification of the Secutypes using the Scala implemented Pattern model allowed for a concise script to be produced by the *aruchi* editor. Scala was chosen for the implementation script for that reason: it provided a concise way of specifying the constraints according to the Secutype model and would be easily used to guide the development of information management systems. The development of these information management systems was required for the projects outlined in section 5.5, which required that the Pattern development use a software tool to parse the Patterns and generate the appropriate user screens and database schema that matched them. This software tool is called the Kiln, which was designed and implemented by the EHR team to emit technical artefacts on the Scala specified Patterns that build the information management tools, for example the Cortext application described in section 5.5.4. The author used this tool to develop the information policy-editing tool, which would ensure that it was developed consistently with the constraints specified in the Secutype Scala Scripts.

The Kiln generated three sets of scripts when it processed the details represented in the Scala scripts: the first was a set of Structured Query Language (SQL) scripts that defined the database schema for the Web Application; the second was a series of Java libraries that defined the business logic for the application, including the functions to store and retrieve data from the database, and the third presents the data in a meaningful form to the user and allow the creation, update and revision of the data. Figure 36 shows the process of authoring Scala scripts, which are then used by the Kiln to generate the SQL scripts for the information management tool database, the Java based business logic and the tooling user screen that were implemented using Java Server Faces.

The information management tools produced by the Kiln established the Clinic Manager 3 (C3) framework, within which the tool components were developed.

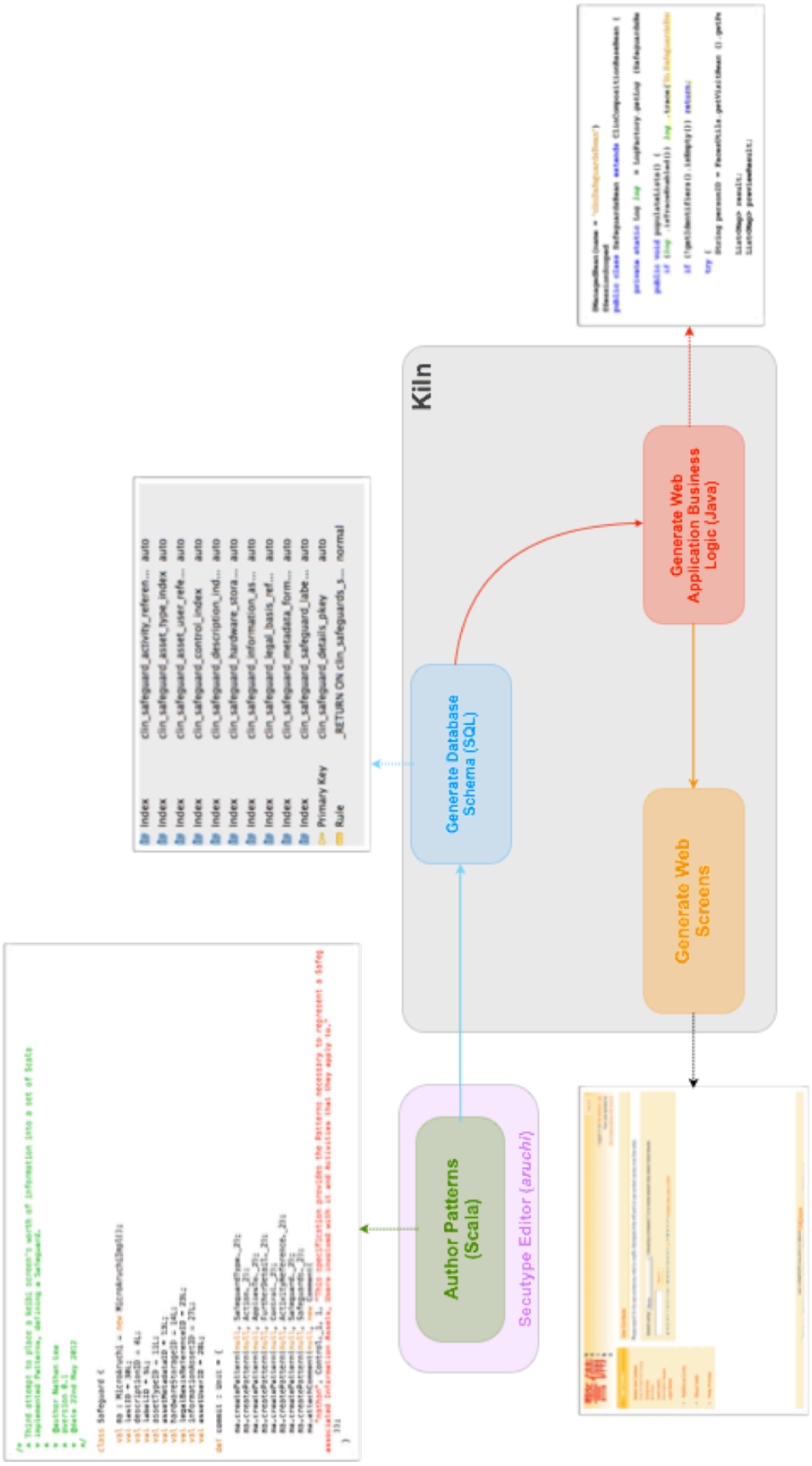


Figure 36: Process for authoring Secutypes and generating policy record tool

This framework was implemented using the Enterprise Java Framework, establishing an n-tier architecture as described in section 4.5, which included a presentation tier developed using Java Server Faces, a business logic tier using Java and a persistence tier using the PostgreSQL database. The J2EE framework was chosen as the deployment infrastructure because of its proven success in implementing EHR systems in the case study examples (described in section 5.5) and evolution from the previous versions of the EHR systems developed for these case studies. The development of a web-based application was a key decision to ensure that users of a policy-authoring tool would have a shared, consistent view and access to information items that had been updated by other users. C3 was also developed to run as an EN 13606 compliant record server, which meant that the framework would have to be able to support versioning of record components as they were added and updated, as well as the logging of access and contribution to the records that it managed.

This made the C3 framework an ideal choice to use for implementing the policy editing tool managed by Secutypes in order to satisfy requirements GEN-13 in Table 2, POLED-2, POLED-3 listed in Table 9 and the audit requirements listed in Table 10 The next section describes the policy editing tool that has been developed and implemented using the Kiln and Clinic Manager 3 framework. The description of this tool provides other details of the features of the C3 framework, which have been essential to meeting the requirements defined in Table 9.

7.5. Policy authoring tool - *keibi*

The use of the Secutype Patterns configures the Kiln to produce a web application, which serves as the tool by which information governance policies are authored, retrieved and reviewed. The tool that has been developed is called *keibi*, the features and functionality of which are described in this section. The tool has been named *keibi* (警備) after the Japanese word for “policing” or “governance.” This section describes the system architecture, features and functions of *keibi*, providing screenshots and an explanation of how the tool works both from the user facing perspective and the core architectural components. It also discusses the choices of implementation and use of the Clinic Manager 3 framework.

7.5.1. *keibi* System Architecture

keibi is a web based application based upon a three tiered model: a component to present screens for user interaction with information, some business logic to process the information and a backend database to store the information and associated meta data. The framework used to run the components of *keibi* is the Enterprise Java toolkit, which includes libraries to develop web pages in the form of Java Server Faces, database connectivity and additional business logic to implement the functionality within the application itself. The J2EE framework comprises the front and middle tiers, and is encapsulated in the Apache Tomcat web server. This component connects to a PostgreSQL database management system, which stores the data itself.

7.5.2. *keibi* Administration

keibi offers several administration features, including adding user accounts, location accounts and the ability to assign particular users and Use Contexts to those location accounts. A user account contains details that can be found across many web systems. These are detailed below for Participant APD_00014 in Figure 37. It is possible to enable or disable user accounts, though user account deletion is not yet possible due to the need to maintain an audit trail of user activity in the system.

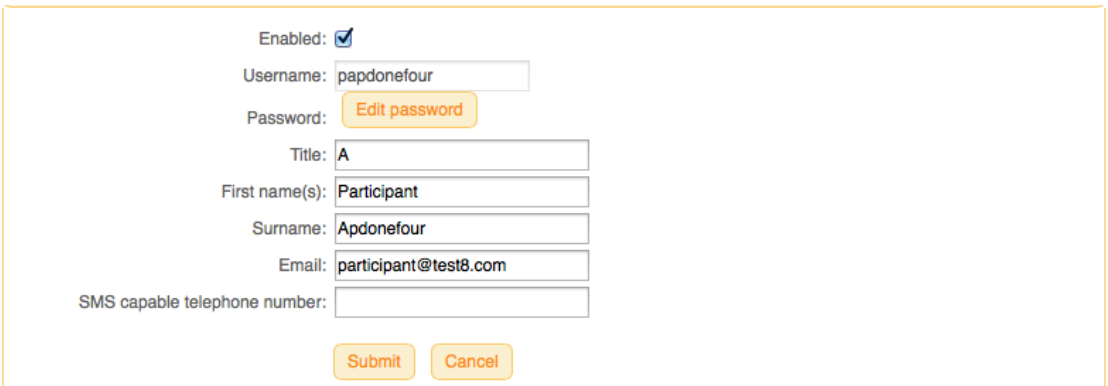


Figure 37: Example user account editing screen

keibi allows the creation of an account defining the location of a particular endeavour. For example, this could be a university site, research institution or clinical site. Users can be assigned to these sites and once assigned, they can

access any Use Contexts stored within them. In the case of the evaluations described in the next section, the evaluation date was defined as the account, and the participants added as users to that account. Figure 38 shows the account details that *keibi* holds using the account developed for the fifth evaluation as an example, which allows it to be enabled, a unique identifier and a description.

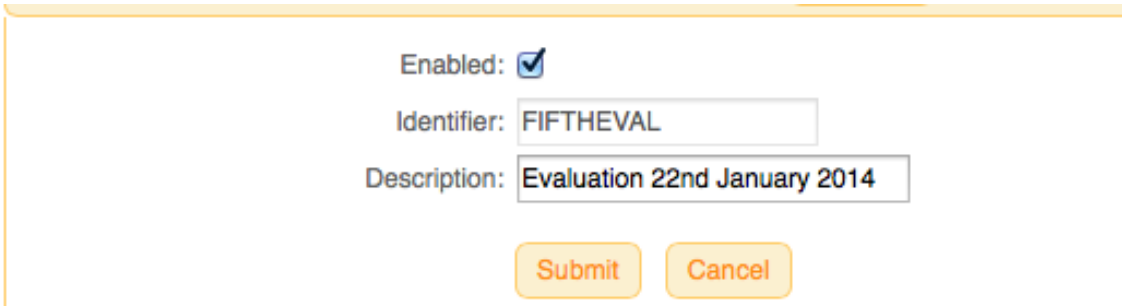


Figure 38: Example of an account editing screen

keibi also has the capability of assigning each user a role linked to a particular sensitivity within a given account. The EN ISO 13606 standards define a set of five sensitivity values, which have been worked into the C3 framework by default and relate to the care context. These roles are linked to specific sensitivity levels as listed in Table 26, which provides access to different record components of an EHR depending on its sensitivity level. These sensitivity values have been defined as system roles within the C3 environment.

CS_SENSITIVITY value	Sensitivity level	Description of intended access to RECORD_COMPONENTs of this sensitivity
Personal care	5	to be shared by the subject of care perhaps with only one or two other people whom they trust most, or only accessible to the subject of care (and to others by one-off authorizations)
Privileged care	4	access restricted to a small group of people caring intimately for the patient, perhaps an immediate care team or senior clinical party (the privileged clinical setting needs to be specified e.g. mental health)
Clinical care	3	default for normal clinical care access (i.e. most clinical staff directly caring for the patient should be able to access nearly all of the EHR)
Clinical management	2	less sensitive RECORD_COMPONENTs, that might need to be accessed by a wider range of personnel not all of whom are actively caring for the patient (e.g. radiology staff)
Care management	1	RECORD_COMPONENTs that might need to be accessed by a wide range of administrative staff to manage the subject of care's access to health services

Table 26: Sensitivity Levels for Role Holders as Defined in EN ISO 13606-4

Table 27 provides the functional roles and a brief description of each of those as defined by EN ISO 133606-4. The standard maps these functional roles to the sensitivity values / system roles as specified in Table 26. This provides the basis for roles based access control and privilege management within the C3 framework.

Functional Role	Brief description
Subject of care	principal data subject of the electronic health record
Subject of care agent	e.g. parent, guardian, carer, or other legal representative
Personal healthcare professional	healthcare professional or professionals with the closest relationship to the patient, often the patient's GP
Privileged healthcare professional	nominated by the subject of care OR nominated by the healthcare facility of care (if there is a nomination by regulation, practice, etc. such as an emergency over-ride)
Healthcare professional	party involved in providing direct care to the patient
Health-related professional	party indirectly involved in patient care, teaching, research, etc.)
Administrator	any other parties supporting service provision to the patient

Table 27: List of Functional Roles from EN ISO 13606-4

	RECORD_COMPONENT sensitivity				
Functional Role	**Care management**	**Clinical management**	**Clinical care**	**Privileged care**	**Personal care**
Subject of care	Y	Y	Y	Y	Y
Subject of care agent	Y	Y	Y	Y	Y
Personal healthcare professional	Y	Y	Y	Y	Y
Privileged healthcare professional	Y	Y	Y	Y+	++
Healthcare professional	Y	Y	Y		
Health-related professional	Y	Y			
Administrator	Y				

Table 28: Mapping of Functional Roles to Sensitivity Values / System Roles from EN ISO 13606-4

The identified requirements for developing *keibi* include one for authorised user access to the system. It is clear that information security policies and governance documentation may have different sensitivity levels across the different components that form part of a particular document, however some elements are intended to be shared widely and publicly. The optimal method for defining access controls and privilege management requirements for healthcare policy has yet to be considered more thoroughly in the area of information governance of healthcare information, but whilst the precise details of how users should get access to different components of a policy document and the kinds of roles and privileges that are used in the context of managing information governance policy are beyond the scope of this work, they are proposed as further work as described in Chapter 10.

The author opted to base the roles, access control and privilege management on the EN ISO 13606 specifications. This was because the policy details could contain identifying, sensitive details from the EHR as the Pepper Potts example provided in section 7.3.5 has shown. Basing the privilege management and authentication on the existing implementation of the standard seemed a sensible approach until a full risk assessment and review could be performed on the use of the knowledge management approach for managing information governance in practice. The discussion in Chapter 9 provides more details in addition to the proposed further work described in Chapter 10.

For the purposes of developing *keibi* and evaluating it in a laboratory setting two of the roles were used: one for an Administrator role, which had responsibility for setting up user accounts, location accounts, granting roles and access and common terms like the Named Properties, the role which was used to provide the care service to users as per EN ISO 13606-4. The other role was equivalent to the Clinical Care role specified in EN ISO 13606-4, which allowed users to access accounts they were authorised to add, edit and review contributions to the policy record for which they had access. The author proposes that Clinical Research would be a more appropriate system role name for the use of *keibi* in the research context. Figure 39 shows the assignment of roles to user accounts and user accounts to accounts.

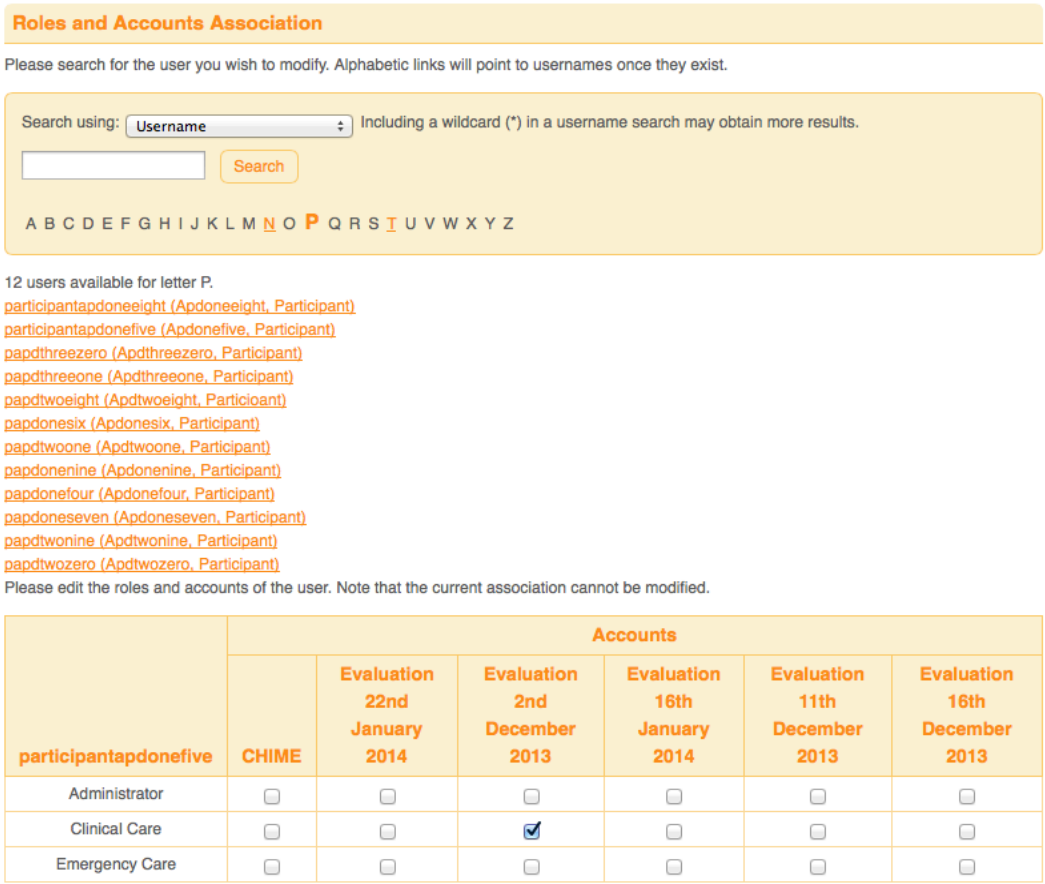


Figure 39: Assignment of roles to a particular user and that user to accounts

keibi also allows the specification of specific properties according to a pre-defined specification in line with the Named Properties data type described in section 7.3. An example list of these properties is shown in Figure 40, which also shows the list of data items associated with that property, in this case, hardware storage. Properties are useful when handling a set of widely agreed terms, or those where users need not be expected to enter the details themselves and can instead pick from a predefined selection. The Administrator Role has responsibility for editing these properties.

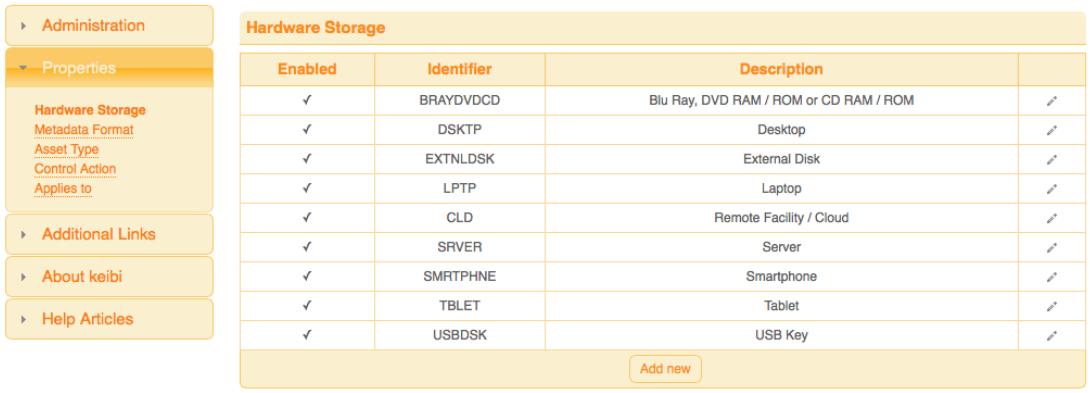


Figure 40: Administration of Properties within the system.

7.5.3. *keibi* Research User screens

On accessing *keibi*, the user is presented with a home screen where they can log in once they have been registered with the system (Figure 41). The list of roles and accounts that have been added to the system are available from the respective drop down boxes. The home page provides access to external links (Figure 42) and an introductory page about *keibi* (Figure 43).

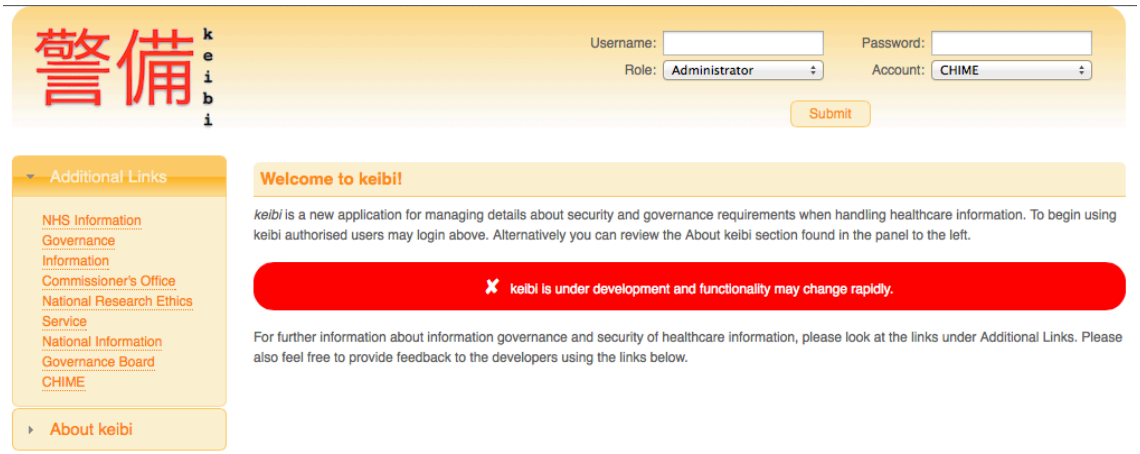


Figure 41: *keibi* home screen

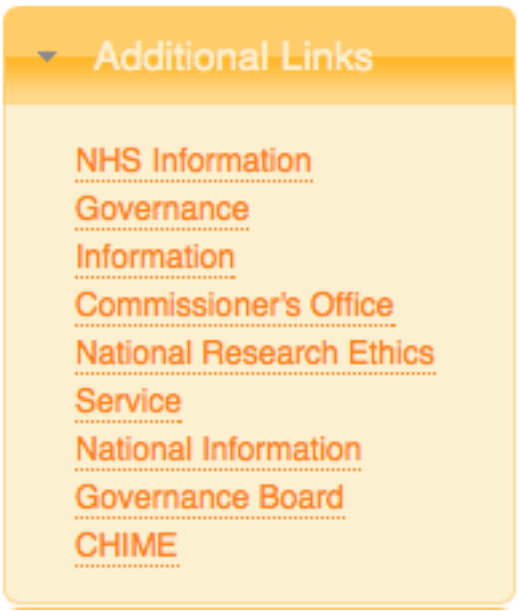


Figure 42: External links from *keibi* home screen

The external links grant access to the associated web page in a new browser window.

Figure 43: *keibi* introduction screen

Once logged in, users have access to more options, to retrieve specific Contexts of Use and the associated policy items that form the Secutype model. Users start by selecting a Use Context from the Use Context Search Screen (Figure 44).

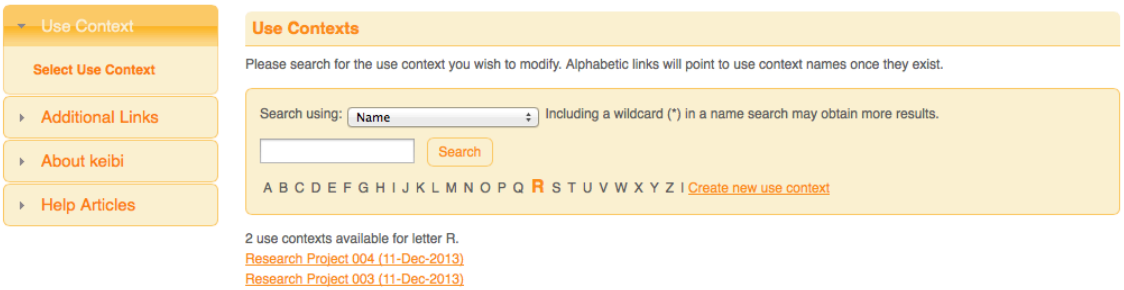


Figure 44: Use Context search and selection screen

Once selected, the user is presented with a summary of the Use Context as shown in Figure 45. Users can select the information governance concepts / Compositions from the list in the Register box, or by selecting them from the left hand drop down menu (Figure 46).

Figure 45: Summary details for a selected Use Context

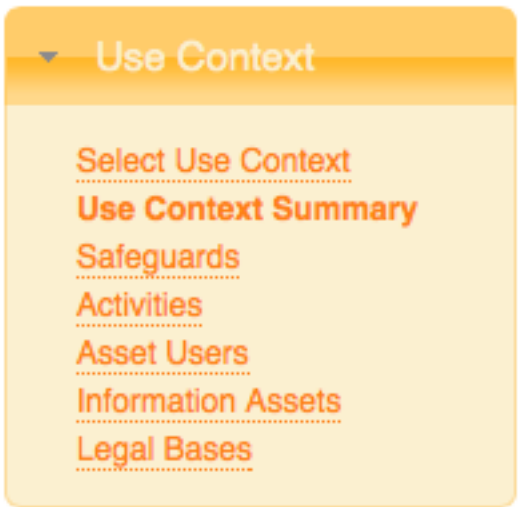


Figure 46: Selection links from the drop down menu, available after selecting Use Context

On selecting one of the concepts, users can create new records or view, add and update the details in existing records. They are presented with a preview of the existing records already stored in the system for each contributed concept, as shown in Figure 47 by way of example for Safeguards.

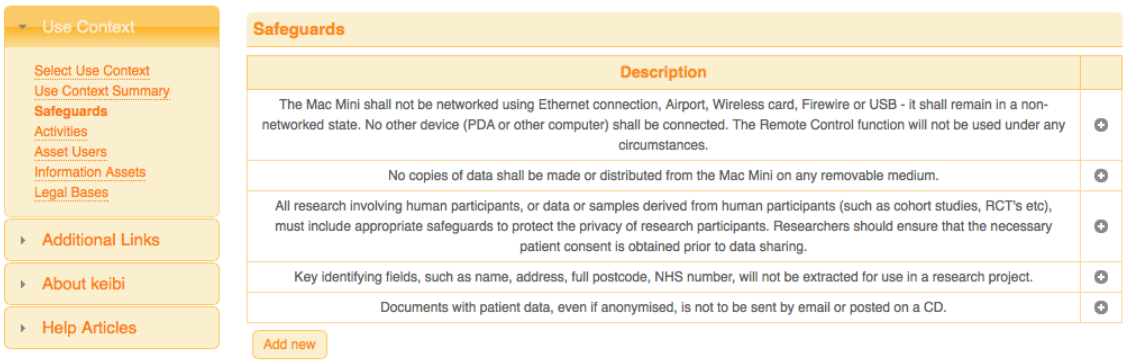


Figure 47: Summary screen of Safeguards added to a particular Use Context

Users may then either view the individual Compositions by clicking on the + symbol, or add a new one, as shown in Figure 48, where each Concept has its own specific screen and details as described in the previous section. Appendix 17 shows the policy excerpts authored by participants themselves in *keibi* during the evaluations.

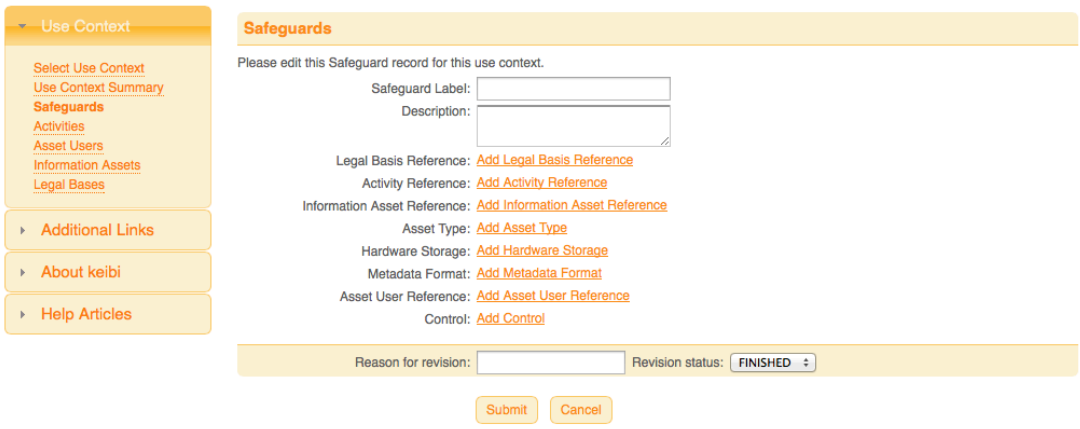


Figure 48: Edit screen for a new Safeguard.

Figure 48 also shows a panel for specifying revision reasons and statuses. This illustrates a feature of *keibi* that is made available by the C3 framework for version management of records. Each time a record is created and then updated, the older version is retained. The system provides this by offering a Revision History button, whereupon users may review older versions of the record and have an

audit history of who created and amended it. This is illustrated in Figure 49, where there are two versions of this particular Safeguard record.

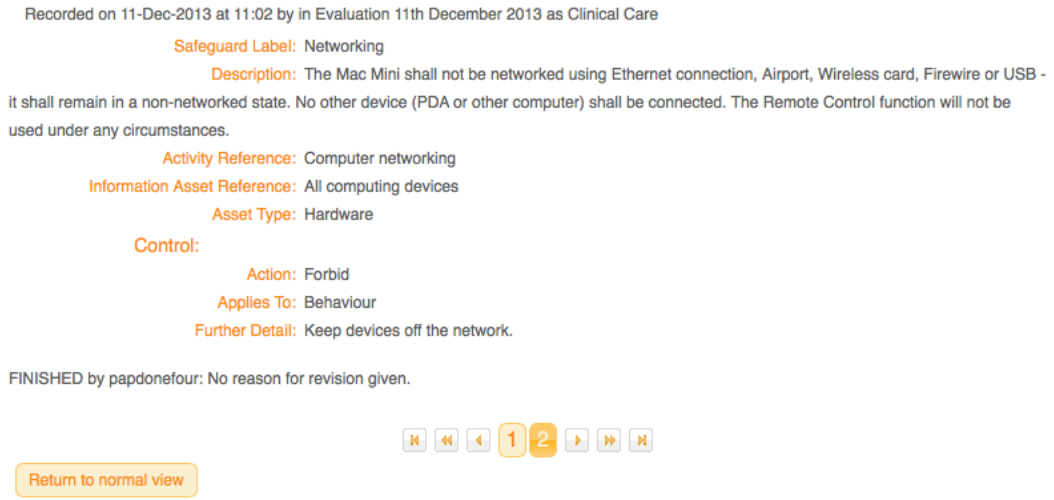


Figure 49: Revision history for an updated Safeguard

keibi also features an extensive range of help screens to assist users should they have any questions or are uncertain about how to proceed using the tool. The list is provided in Figure 50, which shows the titles of the help articles. In total there are seven help articles, and these can be viewed in Appendix 6. They focus on introducing the tool and conventions under *Where do I Start?* and go on to describe each of the six information governance concepts, what they mean and how they can be edited and reviewed.

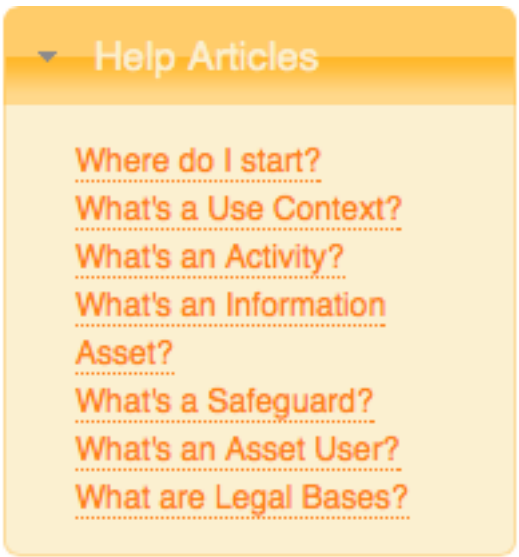


Figure 50: List of help articles available in *keibi*

Another feature of *keibi* is a full audit trail of creation, editing and review of the policy items that are stored and managed by the system, meeting the requirements specified in Tables 9 and 10. The audit feature allows users to view who added, edited or viewed a particular record, when they viewed it and under what role. Figure 51 shows the audit screen to help focus on a particular user or Use Context within a particular date range, in this case for Participant APD_00017, and Figure 52 shows the information provided for the audit.

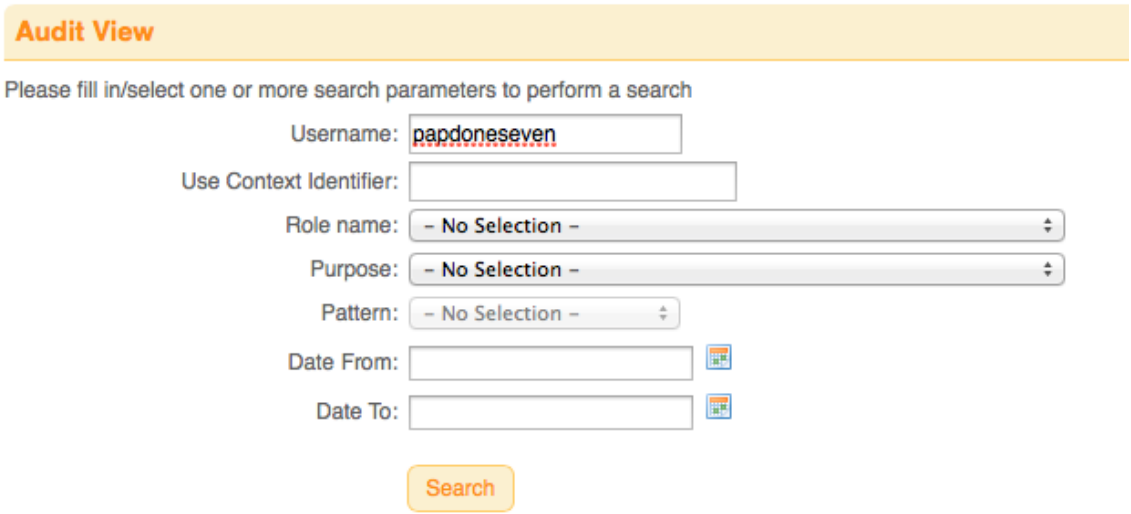


Figure 51: Audit search criteria

There are 199 audits found.

Results 1-20

Date Audited	Actor / Purpose	Use Context Identifier	Pattern / Composition Identifier	Function Name / Notes
16-Dec-2013 15:06	papdoneseven, Clinical Care, Evaluation 16th December 2013 / SUBJECTOFCARE_RETRIEVED	0ca55b2c0ea548a7b46e8e3cc51c7806		{"SELECT FROM subject of care"}
16-Dec-2013 15:03	papdoneseven, Clinical Care, Evaluation 16th December 2013 / AGGREGATE_DATA_RETRIEVED	0ca55b2c0ea548a7b46e8e3cc51c7806		{"SELECT FROM CLIN legalbases legalbasis","SELECT FROM CLIN legalbases legalbasis"} / ClinLegalBasesClinLegalBasis, getIdentifiers()
16-Dec-2013 15:03	papdoneseven, Clinical Care, Evaluation 16th December 2013 / AGGREGATE_DATA_RETRIEVED	0ca55b2c0ea548a7b46e8e3cc51c7806		{"SELECT FROM CLIN legalbases legalbasis","SELECT FROM CLIN legalbases legalbasis"} / ClinLegalBasesClinLegalBasis, getPreview()
16-Dec-2013 15:39	papdoneseven, Clinical Care, Evaluation 16th December 2013 / COMPOSITION_CREATED	0ca55b2c0ea548a7b46e8e3cc51c7806	Safeguards / 92adb9f11a3347f7a6a6476cfd62757b	{"INSERT CLIN safeguards safeguard 92adb9f11a3347f7a6a6476cfd62757b"}
16-Dec-2013 14:57	papdoneseven, Clinical Care, Evaluation 16th December 2013 / AGGREGATE_DATA_RETRIEVED			{"SELECT FROM subject of care"} / Populating alphabet

Figure 52: Excerpt from audit report showing the creation of a safeguard by participant APD_00017

7.5.4. System Testing Using Pilot Evaluation

keibi was developed and implemented by the author using several iterations of the Secutype model specification and continued development of the core C3 framework, including functionality, web page style and layout. The completed package was tested by holding a series of pilot evaluations where other live participants could use the tool according to the proposed thesis evaluation framework described in the next chapter. This process helped to ensure that the system functioned as expected and to gather a preliminary view from pilot users about the development of the knowledge model and what revisions could be made if they found certain aspects unclear or unhelpful. The final evaluation approach is described in section 2.5 and discussed in detail along with the results in the next chapter.

This chapter has provided an overview of the design and development process of the Secutype model and *keibi*, following the Unified Software Development Process. This has included a description of the requirements that have been gathered for the Secutype and *keibi*, use case descriptions and system design using sequence diagrams and the development of class diagrams for the Secutype. This chapter has also provided details of the implementation of the Secutype in Scala using the Pattern constraint model and the implementation of *keibi* using the Kiln tool generation software into the Clinic Manager 3 framework, and the testing of the model and tool in pilot evaluations. The next Chapter describes the main evaluations that were carried out using *keibi* to evaluate the thesis research.

Chapter 8. Thesis Evaluation

This thesis of this research proposed that the knowledge management framework underpinned by the Secutype Classes and implemented as the *keibi* tool clarified the legal, ethical and good practice requirements for people when they specified policy and handled healthcare information for clinical research in accordance with those requirements. It aimed to evaluate the approach by developing the Secutype and *keibi* and assessing them as a proof of concept solution. The evaluations involved testing the following hypotheses, which stated that the knowledge management framework:

1. encouraged a consistent understanding of expected behaviour across a range of role holders when authoring and reviewing information governance policy;
2. limited variation in interpreting information governance requirements when authoring and reviewing policies;
3. supported user expertise when interpreting required behaviour and refined these requirements to computable heuristics.

The purpose of using the knowledge management approach was to clarify and simplify the process of developing effective information governance for people. Human participation in the evaluations was therefore essential to meaningfully test the hypotheses and evaluate the tool as a proof of concept solution. The hypotheses focused on the participants' use of the tool and the evaluation of the knowledge management approach as a proof of concept had to assess how the participants felt using the tool, whether they found it helpful and if they felt it would be of use in practice. A further set of evaluations were held to gather these results, and focussed on answering these questions:

i. how do participants feel about using *keibi*? Do they feel that *keibi* is easy to learn and use?
ii. would participants use it in practice and can they think of a time recently when they would have found the tool helpful?

The evaluations were conducted over three experiments with the involvement of human participants as described in section 2.5. By answering these questions, the evaluations sought to provide evidence of how participants used *keibi*, whether they understood how to use it to author policy items, guide them on how to behave with sensitive information, and whether they felt that it improved the management of information governance. This chapter describes how participants were gathered, and how the evaluations were run, as described in section 2.5. It provides an analysis of the results that were produced from the evaluations to answer the research questions and assess the tool as a proof of concept.

8.1. Participant Responses to Invitation

Participants were asked to provide the excerpts that would be used to develop the first and second experiments once they had accepted the invitation, Three out of the five participants that were approached for the pilot evaluation agreed to participate. In the final evaluations, fourteen potential participants were approached, of which twelve took part in the evaluations. A full description of the participants who agreed to attend the final evaluations can be found in Appendix 11, where participants are listed by the date they attended their evaluation session and are identified by their participant number. The details include an overview of their work areas and responsibilities, and their professional background.

The goal of the evaluations was to include a spread of the different skills and professions that are involved with performing clinical research. Chart 1 shows the spread of participant professions and expertise, where the numbers indicate the number of participants who hold a particular profession or area of expertise. The spread is representative of the skills and expertise that has been found in the research work described in Chapters 3, 4 and 5. Some participants held more than one of these skills.

Participants provided a set of excerpts for both the pilot evaluations and final evaluations, which can be found in Appendix 12. The return rate of excerpts in the pilot evaluations was 100% - of the three participants that attended the session, each provided excerpts. The rate of return for the final evaluation was lower: of the twelve participants that participated, two returned excerpts on time, one of

which included virtually the same excerpts as the other participant. Four others returned them on the day of the evaluation or the day before, which was too late to prepare exercise sheets for the sessions. The other five did not return excerpts due to their not being able to do so in the weeks prior to the evaluation. The author therefore opted to prepare a set of questions that could be used across all the sessions, taking excerpts used during the pilots and from the set provided for the final evaluations. This provided the materials for the exercise sheets used in the evaluation sessions and formed the basis for the questions.

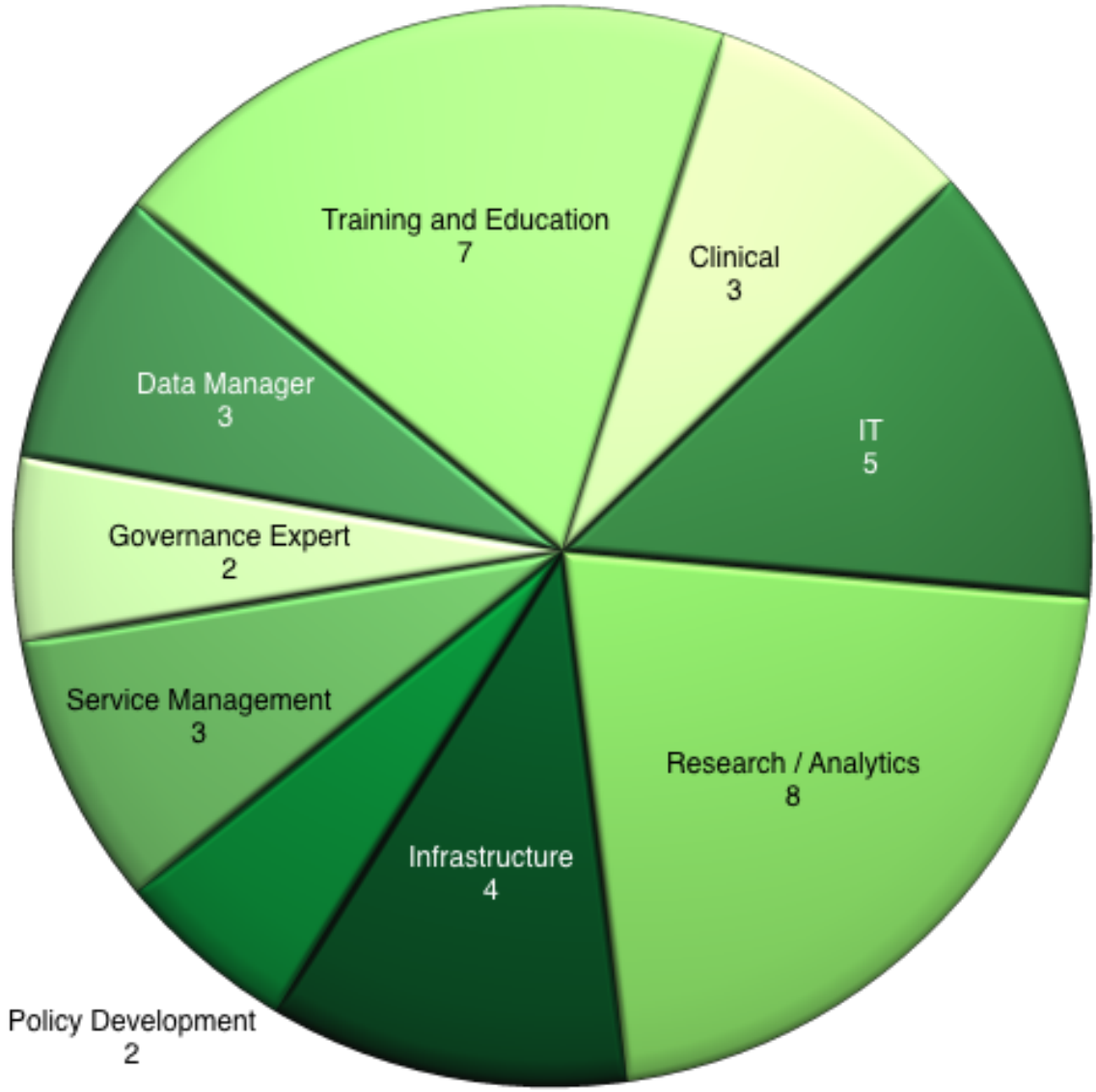


Chart 1: Spread of participant expertise.

8.2. Development of Exercise Sheets

The hypotheses related to the authoring and subsequent use of policy by a variety of role holders involved in clinical research using electronic healthcare records. These exercises provided a basis for gathering participant feedback on using the tool. The evaluations were conducted by using exercise sheets for the participants to answer a series of questions across three experiments. The experiments were designed to allow participants an opportunity to use *keibi*'s features so that they could complete each question and therefore answer the four evaluation questions above. The goal was to ensure that participants worked their way through each of the questions and excerpts, and think about how they would gather the information that they needed to author each element. This would be achieved by authoring them according to the five core information governance concepts that *keibi* presented, encapsulated by the Compositions Asset Users, Information Assets, Activities, Legal Bases and Safeguards as part of the first experiment. Participants would then use the policies authored by their counterparts to answer a series of questions about how they would handle sensitive information in a set of scenarios as part of the second experiment. The evaluations would conclude with a third experiment, which asked participants to fill out a user satisfaction questionnaire and participate in a group discussion to answer questions i. and ii. above. The next section describes the three experiments in detail

There were a total of two exercise sheets used across the twelve participants, containing questions for experiments one and two and the user satisfaction questionnaire used for experiment three. The first four questions for each participant were the same in both exercise sheets, where the last three questions were different to allow the participants to review policy items they had not seen or authored themselves and having no familiarity with the policy items that had been authored by their counterparts in the evaluation sessions. This represented a more realistic scenario for working practice, where policy users would possibly use the policies that they themselves had been involved in authoring, or would perhaps have had no involvement. More participants handled the first answer sheet: this was not by design: where there were a maximum of three participants

in each session and the decision to provide two participants with an adapted Sheet One or Sheet Two was entirely independent between evaluation sessions. The exercise sheets can be found in Appendix 13 and a version that describes the exercise sheets and the expected answers are provided in Appendix 14.

A set of expected answers were developed by reviewing each of the policy excerpts and identifying where there were potential instances of each of the Compositions, which are provided in Appendix 14. This provided a framework to assess participant responses, where assessment of experiment one would also include determining whether participants had understood what they needed to enter, and if they had not, if this was due to their error or if the tool itself misled them. It was of course entirely possible that participants would exceed expectations and provide responses that were not as expected but no less valid, and would provide evidence that their expertise was being supported and guided by the tool, so this too was assessed. The results from experiment two would also be key in terms of assessing whether the participants had authored items that could be understood by their counterparts consistently and with limited variation in interpretation, or whether the tool itself caused ambiguity and misunderstanding. The assessment of the responses to experiments one and two is described in more detail in the results and analyses section 8.6.

8.3. Experiments

The experiments were conducted by means of the exercise sheets described in section 8.2, which provided participants an opportunity to provide more details about themselves, specifically their professional background and involvement with policy development and / or experience of handling sensitive healthcare information. The exercise sheets also included two experiments and the user satisfaction questionnaire, which can be found in Appendix 16: the first asked participants to author the policy excerpts that had been provided either by themselves or their fellow participants. Figure 53 provides an overview of how the experiments were structured during the evaluation sessions, including the introduction and breaks. The following section describes the evaluation sessions.

8.3.1. Experiment One

Experiment one involved participants reviewing the supplied policy excerpts and authoring them in *keibi*. Due to the limited return of excerpts during stage one the author decided to base the questions on the excerpts gathered during the pilot evaluation and the ones supplied during the more recent round of invitations. This provided a more consistent set of questions and a stronger basis for comparison between participants across evaluation sessions. It was also not a reasonable assumption that policy authors would be authoring policies that they had already developed, or had had any involvement in developing; it was therefore a more reasonable test to have them author policy items that they would not necessarily have seen before or otherwise be familiar with.

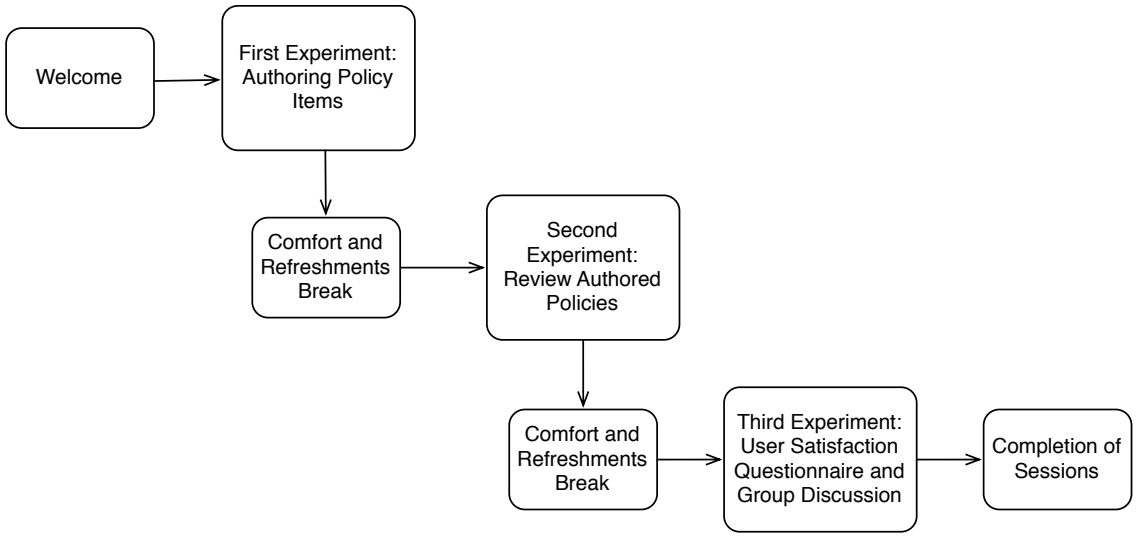


Figure 53: Evaluation Session Experiment Flow

The experiment took approximately one hour. There were a total of seven questions, the first four of which involved the same tasks and use of the same excerpts across the participants and were treated as a warm up exercise for participants to get used to handling *keibi*. The first question asked participants to enter a CD ROM Information Asset, the second, a Legal Basis in the form of Patient Consent. The third asked participants to enter a Safeguard, which also provided a policy item that was intended for software configuration as well as human readership. The fourth question also expected participants to specify a Safeguard and this was timed. These were all based on excerpts that had been supplied

during the pilot evaluation. The fifth, sixth and seventh questions were based upon excerpts provided by some participants in the recent round of invitations as described under Stage One as illustrated in Figure 2.

Questions three, four, five, six and seven left the participants to decide on what should be entered on *keibi*: a series of expected details were developed so that the details that participants entered could be compared with these expected outcomes. It would therefore be possible to determine how intuitive the participants found *keibi* and the Knowledge Management Framework it incorporated to enter details and use. This would be correlated with how participants answered the questions in experiment two, which is described below, to determine how well participants understood what they were authoring and whether their counterparts could answer questions correctly based on what they had entered.

8.3.2. Experiment Two

Experiment two took approximately half an hour and asked participants to specify how they would handle a series of information handling tasks based on the policies that were available in *keibi*. Participants accessed the policy details in Use Contexts that were provided by their fellow participants to author their policy excerpts so that they would review details that they themselves had not entered. The first of the questions was the same across each participant to see how each interpreted the policy requirement and whether there were any differences in authoring that caused variation in results based on the same original policy excerpts that participants had authored. The other questions were different across participants, depending on the excerpts that had been submitted and entered on *keibi* so that they could use them meaningfully to answer the questions; this had the additional benefit that participants could not copy each others' answers. One of the questions was multiple choice, providing the options for participants to select one of a series of answers. In all cases, participants were asked to specify how they had come to answer the questions as they had done.

8.3.3. Experiment Three

Participants were asked to complete a User Satisfaction Questionnaire on completing the first two experiments, which is provided in Appendix 16. This questionnaire has been reused from a set of experiments as described by Lewis (James R. Lewis, 1993) because it represents a well validated set of questionnaires that have been used for evaluating user experiences with software and IT systems. This questionnaire also allowed participants to explain their answers in more detail. The questionnaire chosen was the Post-Study System Usability Questionnaire, which represents an optimal questionnaire for evaluating user experience and usability after using software systems in a lab setting as part of an evaluation.

Experiment three also involved a group discussion over lunch, which offered participants an opportunity to answer some questions as a group and provide a more detailed account of their opinions and experiences beyond their questionnaire responses. The group discussion took about half an hour and allowed more in depth feedback about *keibi*, its use, features, whether the participants would find it useful in their work and whether there has been an occasion in the last year when they would have found it useful. The intention was to supplement the satisfaction questionnaires, thereby gaining insight into their experience and to reflect on their responses and the outcomes, and take any criticisms or suggestions to improve the *keibi* use experience and effectiveness. The questions posed were:

1. What are participant views on the system when they discuss as a group?
2. Would they use it in their working practice?
3. Can they think of a time I the last year when they might have used it?
4. Did the tool make participants think about issues surrounding information security and governance?

8.4. Evaluation Sessions

The evaluation sessions were held on 2nd, 11th and 16th December 2013, and 16th and 22nd January 2014. Each session was delivered consistently, lasted up to

three hours and had no more than three participants per session. Refreshments and lunch were provided along with breaks between the experiments to ensure that the participants were not fatigued by a lengthy experiment process and remained comfortable so that they could focus on the experiments and responses they provided. Participant numbers per session were capped at a maximum of three so that the investigator could provide reasonable attention to each group, and a minimum of two so that each participant could refer to what they had each authored. Table 29 shows the dates when each participant attended and the exercise sheets they were provided during each of the evaluation sessions. Participants had no prior knowledge of who would be attending the evaluation sessions with them but may have worked with fellow participants before. This represented a fair representation of actual working practice as it would not be safe to assume that potential users of *keibi* would either know their colleagues prior to working together or would never have collaborated before.

Participant ID	Date of Attendance	Exercise Sheet Provided
APD_00015	02/12/2013	Sheet 1
APD_00018	02/12/2013	Sheet 2
APD_00014	11/12/2013	Sheet 2
APD_00019	11/12/2013	Sheet 1
APD_00016	16/12/2013	Sheet 2
APD_00017	16/12/2013	Sheet 1
APD_00021	16/12/2013	Sheet 1
APD_00029	16/01/2014	Sheet 2
APD_00030	16/01/2014	Sheet 1
APD_00020	22/01/2014	Sheet 2
APD_00028	22/01/2014	Sheet 1
APD_00031	22/01/2014	Sheet 1

Table 29: Dates Attended by each participant and exercise sheet tackled

The evaluations were held in a meeting room in the Centre for Health Informatics and Multiprofessional Education, where there was plenty of room for participants, and the room temperature was comfortable. Apple MacBook Pro computers were used to conduct the evaluations, except in one case where a participant brought their own laptop. Participants used the Apple Safari browser, Google Chrome or Mozilla Firefox. Participants were all familiar with the laptops and browsers

provided. The investigator opted to keep the experiment on a local network and independent of UCL's corporate infrastructure to ensure that any issues with these facilities did not interrupt the experiments.

The session started with a brief introduction to *keibi* and discussion about its functionality to make sure that participants understood the core concepts of the tool, where the focus was on providing a brief introduction to Asset Users, Legal Bases, Information Assets, Activities and Safeguards and Contexts of Use. The introductory presentation also gave participants an opportunity to see if they had any points, questions or feedback prior to using the tool. The introductory slides are included in Appendix 15. The introduction was deliberately brief to ensure that participants had minimal preparation or guidance: this was essential to ensure a fair and rigorous test of how intuitive *keibi* and the knowledge management approach were for participants to follow, and whether participants understood the use the five information governance domain concepts represented by the Compositions. Once the tool had been introduced, the experiments were started.

8.5. Method of Analysis

The results gathered from each experiment captured a range of data about the use of *keibi* to specify information governance policies and advise users on how to handle information processing activities. This section describes how the participant responses for the three experiments have been analysed to evaluate the three hypotheses and the use of *keibi* as a proof of concept solution.

In determining metrics for assessing participant responses in the three experiments, the focus of the first two experiments was on determining how participants used the Secutype knowledge model that was presented to them by *keibi*. The analyses focussed on the extent to which the use of the tool helped to clarify the authoring requirements for policy items so that the subsequent use of those items was not prone to uncertain and inconsistent interpretation and understanding of the principles upon which they were based. This relied on assessing whether they used *keibi* to author the policy excerpts according to expectations described in the previous chapter, whether *keibi* assisted their understanding of how to author and use policies, whether the authoring and

interpretation was achieved consistently, if the details they entered were correct according to what the policy excerpts were specifying subsequently and whether they were correctly used to inform their decisions in the second experiment. Participant opinions on using the solution were key to assessing the effects of the proposed solution and the third experiment was designed to capture as much detail as possible from them about their experience and how it helped their understanding and interpretation of what was required so the evaluations included a satisfaction questionnaire and focus group discussion.

The first hypothesis was tested by assessing whether participants had demonstrated an understanding of authoring policies or misunderstood what they were doing during experiment one. This was measured by assessing whether they were able to author the excerpts and how they used the knowledge model classes to author and them in *keibi* based on the excerpts that were provided in the exercise sheets. These were compared against evidence of misunderstanding, which was measured by assessing errors or omissions in authoring. The first hypothesis was also tested by assessing the responses from experiment two, where participants demonstrated understanding by answering a series of questions consistently with a set of predetermined, expected answers. Consistency in understanding has been determined by looking for trends in the number of cases of identified understanding and misunderstanding and any variations in those numbers. A higher ratio of understanding to misunderstanding would show that using *keibi* encouraged understanding, whilst a consistent number of cases of understanding would demonstrate a consistent understanding. These would both support the first hypothesis.

The second hypothesis was tested by analysing the number of different Secutype model Compositions that participants had added for authoring the policy excerpts in experiment one and scoring their responses. Assessing the numbers of responses from experiment two also tested the second hypothesis. A consistency in the different Compositions used to answer the questions and the resulting scores would support the second hypothesis by demonstrating that participants had a limited variation in interpreting how to author and then use policy items in *keibi* based on their responses.

The third hypothesis was tested by identifying and assessing cases where participants had used their own expertise correctly and had exceeded expectations when answering the questions in both the first and second experiments. The fourth question in the exercise sheets also provided a computable policy item where the addition of an appropriate Safeguard would provide the basis for a policy item that could be refined to a computable specification. This would show evidence that *keibi* had encouraged behaviour that was consistent with good working practice, and empowered participants to exceed expectations and act correctly where the policy developers and the author had not been able to predict outcomes of using those policy excerpts. Sections 8.5.2 and 8.5.3 describe how understanding, misunderstanding, exceeding expectations and scoring were measured to gather the results used for the analyses.

The first two experiments were developed to test the hypotheses and provide evidence that the tool fostered understanding, supported people in developing information governance guidelines and assisted their working practice. Gathering data on participant feedback about their use of the tool was necessary to provide a more complete evidence base for the effectiveness of the tool as a proof of concept solution. The third experiment was therefore run to develop a further evidence base for assessing the tool as a proof of concept solution. The evidence was gathered by collecting responses to the user satisfaction questionnaire and conducting group discussions after the questionnaires had been completed. The results of the questionnaires are provided in section 8.7 along with the group discussions, which were reviewed for common themes and are listed in **Error! Reference source not found.**.

8.5.1. Understanding and Surpassing Expectations

For the first experiment, reconciling the participant responses with the expected outcomes was key to determining whether participants had understood the policies that they authored and how to author them. If participants had misunderstood, it was important to determine the reason for this misunderstanding: this might have been because the participant had misunderstood the original policy excerpt, had made an error using the tool or had

been somehow misled by the tool itself. The analysis therefore notes where participants had demonstrated understanding and misunderstanding, along with the reason behind the misunderstanding. Each case of demonstrating understanding and misunderstanding was counted.

The author recognised that the expected responses may not have covered all possible responses or permutations for answering a given question: the analyses have therefore noted where participants have performed beyond expectation, particularly since this would provide further evidence that the tool supported user expertise. In some cases, this would be by adding extra detail, or by providing a simpler response than expected. This was counted according to each case where expectations were surpassed. The same approach for counting was applied for both experiments one and two for assessment, regardless of whether participants had surpassed expectations.

For Experiment Two, understanding was determined and counted based on whether answers provided were consistent with the expected responses based on understanding the materials entered into *keibi*. Where a misunderstanding was apparent, it was noted as to whether this was due to participant error, authoring error or whether the use of *keibi* had misled the participant into making the incorrect response. The use of participants' own expertise and exceeding of expectations were also counted when they occurred, as well as cases where participants omitted responses.

8.5.2. Correct, Incorrect Response and Omission Scoring

Participants may have understood authoring excerpts in *keibi* but they may have made small errors that were noted and explored: for the first experiment, a scoring system based on the EN ISO 13606 Reference Model has therefore been developed. Responses were scored according to details that had been entered in each Composition: Each Composition was scored at five (5), Cluster at two (2) and each Element has been scored as one (1). The score of five for each composition would be additional to each Cluster and / or Element it contained. A Cluster was scored in addition to the number of Elements it contained. If a participant entered something correctly, they received a *correct response* score, and an *incorrect*

response score if they made an error. Where Participants had omitted key details, the same scoring rules were applied based on what had been omitted (so an entire Composition omission would be scored at five based on the expected response and the constituent Elements and Clusters).

Compositions were scored at five due to their being an encapsulating component in the information governance record hierarchy. They represent the parent node, which frame each of the five information governance concepts and contain the other record components that are represented by Clusters and Elements. Their inclusion, correctly or otherwise, and omission would represent a significant part of determining the accuracy with which participants would enter the details from the policy excerpts. Clusters were scored two because they represented Controls and a collection of Elements meaning that their inclusion or omission were still more significant than a single, standalone element and core to the representation of the policies within *keibi*. Elements themselves were scored as one – whilst they held the actual data values to represent the information governance concepts, these existed within the concepts represented by the Compositions and Clusters, and their individual values held less significance than the encapsulating classes. A higher score would disproportionately represent them given that the concepts that they were describing were being fully represented by the Composition and Cluster that contained them and represented them in a meaningful fashion. Correct or incorrect use or omissions would be less critical than the inclusion of the containing Composition or Cluster as a whole.

For the second experiment, scores were applied based upon whether participants answered the questions provided, where they received a score of one for each correct answer, mistake or omission. The mistake scoring was based on the reason for the mistake: it might have been due to a mistake on the part of the participant interpreting the details in *keibi*, as a result of an error being made in the authoring of the policy in the first place, or due to a failure on the part of the tool to adequately express the rules. Omissions were categorised in the same way. Participants applying their own expertise to answering the questions were also recorded: whilst the capabilities of participants were not being assessed in these experiments, this use of their own experience was also scored where a participant

had indicated that they had done so or it was evident from their responses independently from the expectations.

8.5.3. Reconciliation Between Experiments

There was a clear relationship between each of the experiments: the first experiment informed the second and the third provided an opportunity for participants to reflect and provide their own views on what they had experienced with the previous two. Since the first directly informed the second, any authored excerpts would directly influence how participants answered the questions in the second; for example, errors made in the first might have affected participant responses in the second, leading them to either answer incorrectly, or correctly if the nature of the error or misunderstanding in authoring the excerpt was such that it did not undermine the intended outcome of using the original excerpt.

The third experiment gave participants the opportunity to reflect and discuss their use of *keibi* in the evaluations. They would refer to their experience of using the tool and provide further insights into their responses and the reasons behind them. This has been indicated in the analysis for each question when it occurred, for example when a participant shared that they had used their own expertise instead of the policy details that had been entered into *keibi* by other participants.

8.6. Evaluation Results and Analyses

The evaluation workshops concluded successfully and captured a range of measures from the three experiments. This section presents the analyses of the results from these experiments. The analyses have been conducted to test the hypotheses and summarise the participants' views in their experience of using the tool so that the use of *keibi* as a proof of concept tool could be validated. This section presents the aggregated analyses according to the results of the validation of the three hypotheses and participant feedback. The chapter concludes with an overall discussion of the results gathered form the three experiments.

All answer sheets have been transcribed and had the results collated. The transcriptions and results generated from the answer sheets can be found in Appendix 17 for all experiments; Appendix 18 provides the analysis of the results

prior to aggregation for experiment one and is categorised by each question. Appendix 19 provides the results categorised by each participant for experiment two. A transcription of the group feedback sessions is provided in Appendix 20, categorised by the question posed as part of the discussion in these sessions.

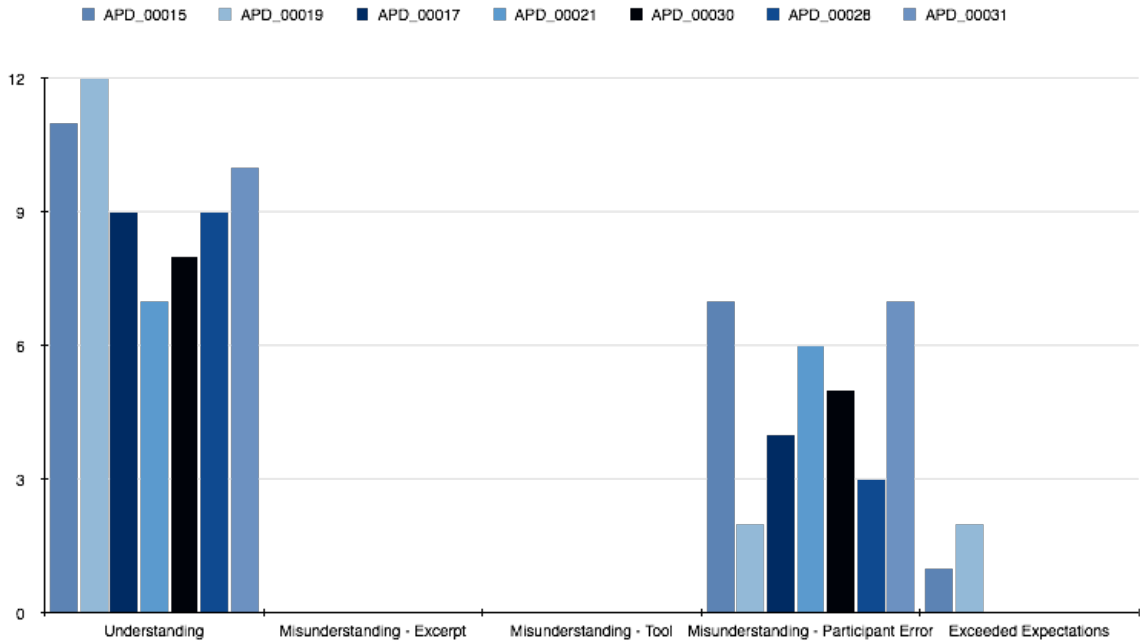
8.6.1. Results of Testing Hypothesis 1

The first hypothesis proposed that the knowledge management framework encouraged a consistent understanding of expected behaviour across a range of role holders when authoring and reviewing information governance policy. The first and second experiments provided results that showed measures of understanding across each of the participants for each of the questions. These measures showed in the first experiment whether participants were able to understand the policies they were authoring according to an expected response or if they exceeded expectations, or if they misunderstood what they were authoring and the reason for that misunderstanding, which could be due to their misunderstanding the original excerpt, misunderstanding because the tool mislead them or misunderstanding as a result of their own error.

The total number for each of these measures also indicates the levels of consistency in understanding when authoring policies across each participant. For the second experiment, a measure of whether they entered a correct response, exceeded expectations, applied their own expertise or made a mistake or omitted expected answers due to the tool misleading the participants, due to an error on the part of the authoring participant, or due to an error by the participant themselves was used. The total numbers across each of these measures would also provide an indication of the consistency of understanding and interpreting the authored policies across each of the participants.

By aggregating the measures of understanding for the first experiment across all participants for all questions, Graph 1 shows the aggregated levels of understanding displayed by those who tackled Exercise Sheet One for the first experiment and Graph 2 shows these levels for the first experiment for Exercise Sheet 2. From these graphs it is possible to see the overall levels of understanding against misunderstanding and the exceeding of expectations across the authoring

of policies, which showed whether the levels of understanding showed a consistency of understanding across the participants for all of the questions.



Graph 1: Aggregated Results for Understanding, Misunderstanding and Exceeding Expectations for All Participants and Questions Across Exercise Sheet 1

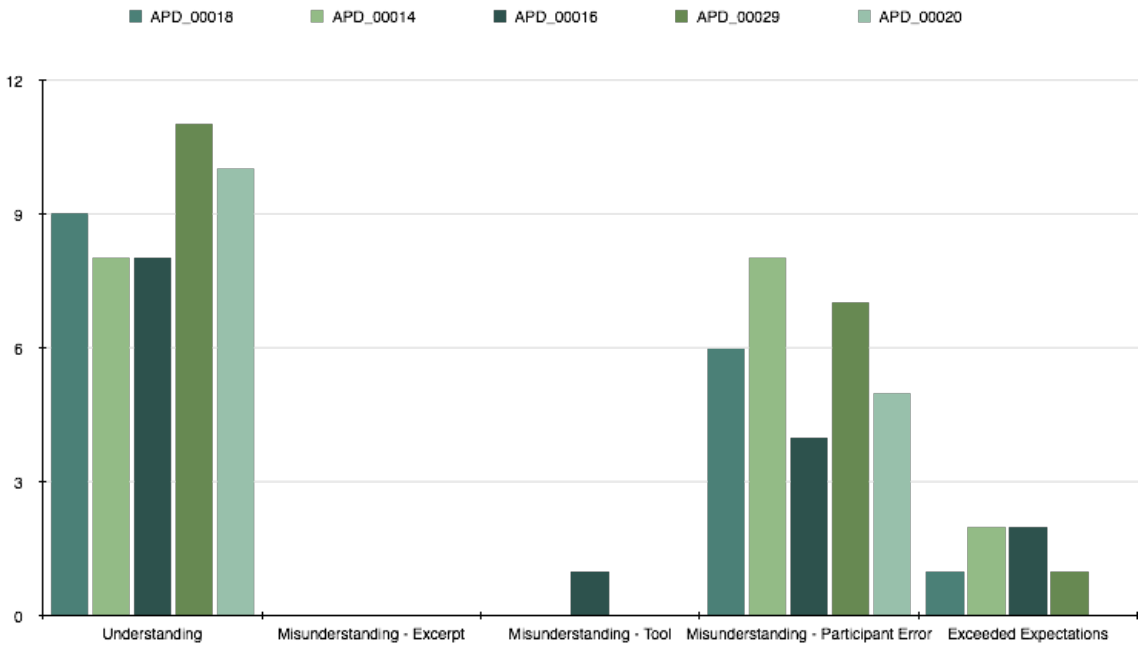
Graph 1 shows the aggregated results across all participants who attempted exercise sheet one for the policy authoring exercise. These results showed that *keibi* encouraged understanding over misunderstanding across the participants, where there was a higher proportion of understanding of policy authoring to misunderstanding, and some evidence of participants exceeding expectations. It also showed that the maximum number of cases of understanding was twelve, and the minimum was seven. There is some variation between the levels of understanding across the participants, which does not support the hypothesis that the knowledge management framework encourages a consistent understanding for these participants when authoring policy.

The proportions of misunderstanding to understanding were however quite high in some cases, where only Participant APD_00019 had only two examples of misunderstanding against twelve examples of understanding and the variance between and the others had a lower difference between understanding and misunderstanding because of participant error. This trend shows that participant understanding was supported and encouraged, but some level of

misunderstanding remained. In the case of APD_00021, who had a lower level of understanding than the others, this participant made a mistake using the tool that resulted in an entire Safeguard was lost without it being saved; the participant was frustrated by this and it affected their performance for entering the policy details given the time it took them to recover from the error and the ensuing frustration. This situation is discussed in the analysis of Graph 56 on page 620.

The aggregated results show that using *keibi* when authoring policy encouraged more understanding than misunderstanding. There was clearly some variation in the understanding when authoring the policy. There was also a higher level of misunderstanding than desirable. This tends to refute the hypothesis that *keibi* encouraged a consistent understanding between participants who attempted this exercise sheet has been refuted.

Graph 2 shows the aggregation of understanding results for writing policy across all participants who answered Answer Sheet 2.



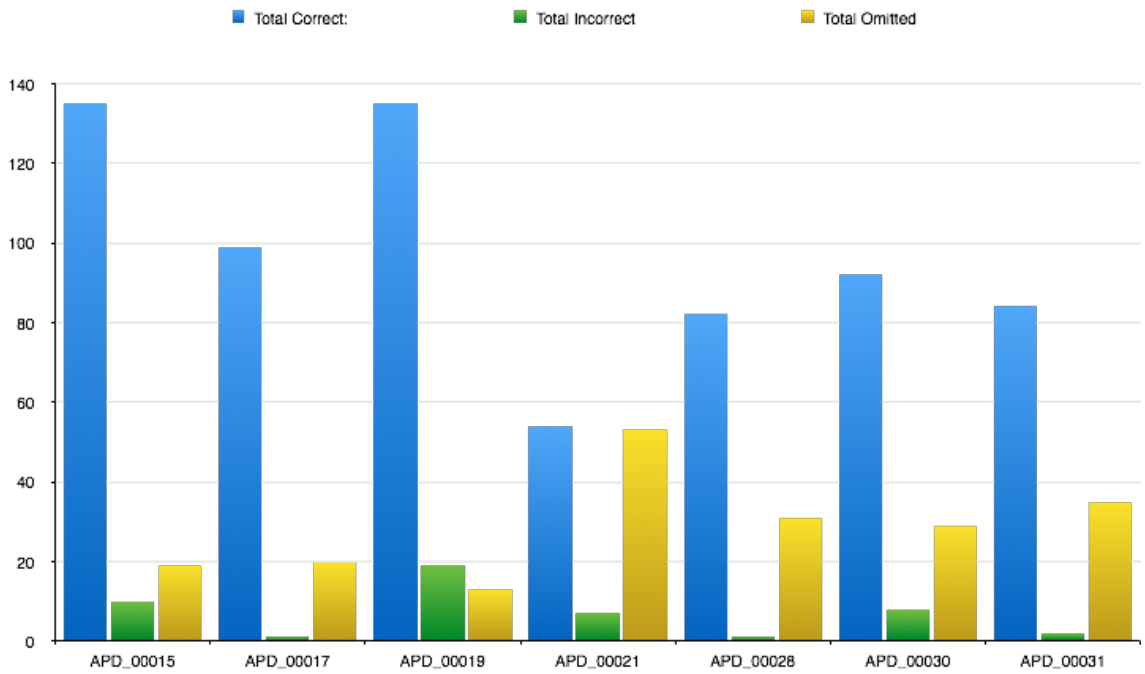
Graph 2: Aggregated Results for Understanding, Misunderstanding and Exceeding Expectations for All Participants and Questions Across Exercise Sheet 2

A similar trend to the results of Answer Sheet 1 can be seen here. There was some variation in understanding across the levels of understanding, without the outlying result for participant APD_00021. There were fewer examples of participant misunderstanding than understanding, but the proportion remained high. There

was also an example of misunderstanding introduced by the tool itself, which is the only example across the two experiments.

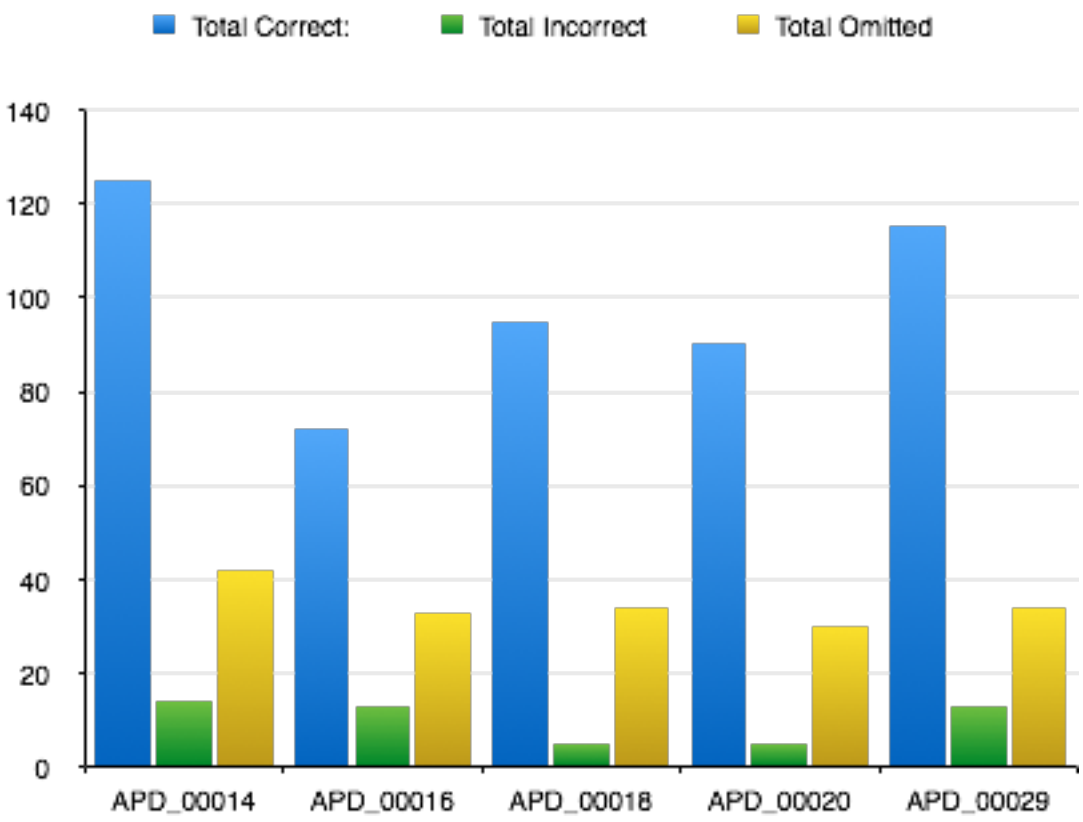
There is more evidence of participants exceeding expectations than in the first answer sheet across the participants, where there was at least one example of this for four of the five participants, showing some consistency across the responses to Answer Sheet 2. This is not consistent with the numbers for the first answer sheet, which may be as a result of some variation in the complexity of the questions and / or interpretation of what was required. In either case, this does not support the hypothesis that the knowledge management framework encouraged a consistent understanding of authoring policy for the participants who answered Exercise Sheet 2. This is compounded by the consistency of the outcomes for understanding and misunderstanding for responses to Exercise Sheets 1 and 2.

The analyses have also scored the responses so that the analysis of the levels of understanding against misunderstanding could take into account the nature of the responses and whether they were more critical or less critical, showing more clearly how participants understood or misunderstood the elements of using the knowledge model to author policy. This relied on scoring the authored entries to see how well participants had performed and the numbers of Compositions, Clusters and Elements that had been added. As described in section 8.5, the scoring was weighted according to the Class that had been entered by the participants in *keibi*. Compositions were weighted the highest at five given that they were the containing class and the overall representation of information governance requirements, where Clusters were scored at two and Elements at one, representing lower significance individually than the overall presentation within the Compositions themselves.. By analysing the scores and counting the variation of numbers of correct, incorrect and omitted responses, the author could assess how significant the misunderstanding was to the overall specification of each policy item and make further inferences about how *keibi* encouraged understanding and consistency in that understanding across the participants. Graph 3 and Graph 4 show the total scores for each participant across Exercise Sheets One and Two respectively.



Graph 3: Total Scores For Exercise Sheet 1

Graph 3 shows the aggregated scores across all participants who tackled the first exercise sheet. This reflects how the participants fared based upon the details that they entered into *keibi* during the first experiment. With the exception of Participant APD_00021, there appears to be a much higher score for correct entries when compared to incorrect entries. Though the proportion of omissions was higher than errors, this still remains low. These results show that *keibi* encouraged correct authoring and that the errors that came about due to misunderstanding were mainly as a result of entering incorrect Elements or omitting them. This trend can be seen in Graph 24 to Graph 67, found on pages 595 to 630, where a breakdown of the number of correct, incorrect and omitted Elements can be found and compared with the Compositions and Clusters. This provides evidence that the majority of misunderstandings shown in Graph 1 and Graph 2 were less critical to the overall development of policy, where they related mainly to incorrect or omitted individual Elements and not entire Compositions. There were higher levels of omissions than incorrect Compositions, and these related to higher levels of both Elements and Clusters. The only Clusters that were available were Controls within Safeguards, and the results show that there were several cases of these being omitted.



Graph 4: Total Scores for Exercise Sheet 2

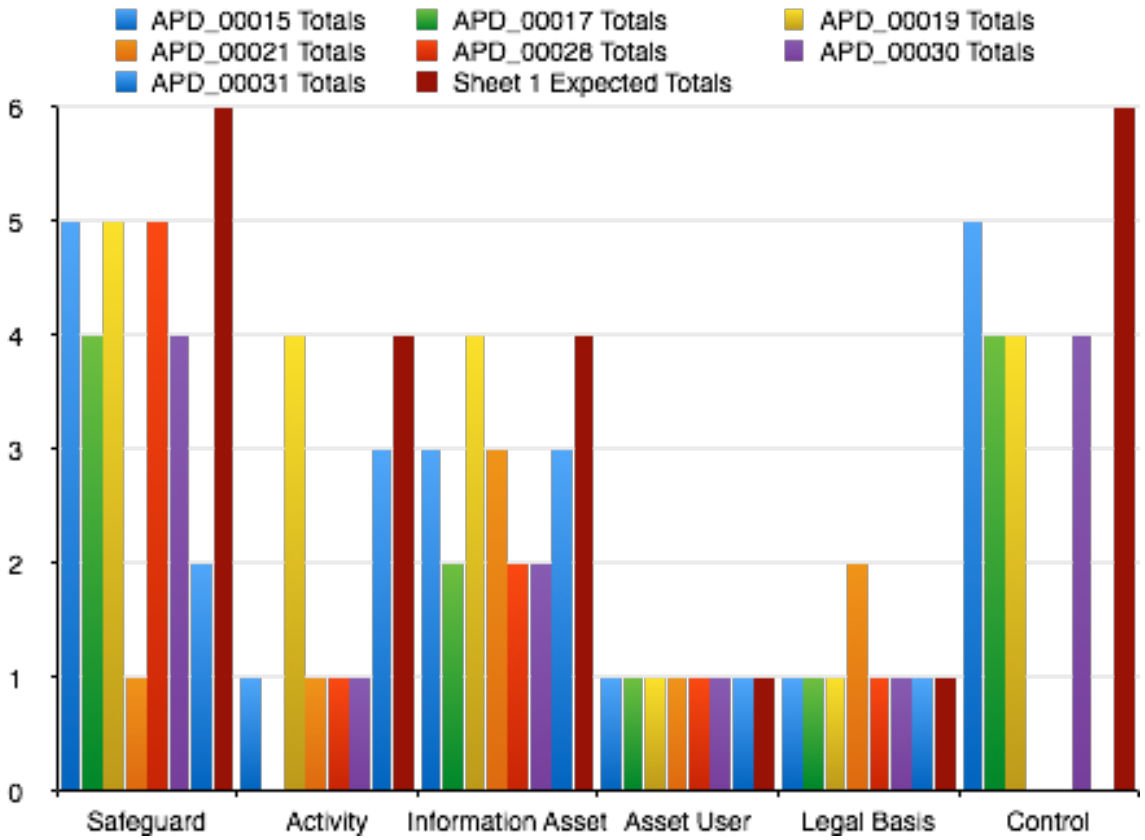
Graph 4 shows the total scores for participants who attempted Exercise Sheet Two. These results are consistent with the results for Exercise Sheet 1, where there was a higher proportion of correct scores to incorrect and omission scores. Omission scores were higher than incorrect scores. The omissions included mostly both Elements and Clusters, with some Compositions, and the errors were mainly due to incorrect individual Elements.

In both exercise sheets there is still some variation in the scores for correct entries, which tends to refute the first hypothesis in terms of providing a consistent understanding across a range of role holders. They do however show that the tool promoted understanding and correct responses, where incorrect responses and omissions were comparatively minor.

8.6.2. Results of Testing Hypothesis 2

The second hypothesis proposed that the knowledge management framework would limit variation in interpreting how to author and use information governance policies. The first experiment focussed on the writing of policies

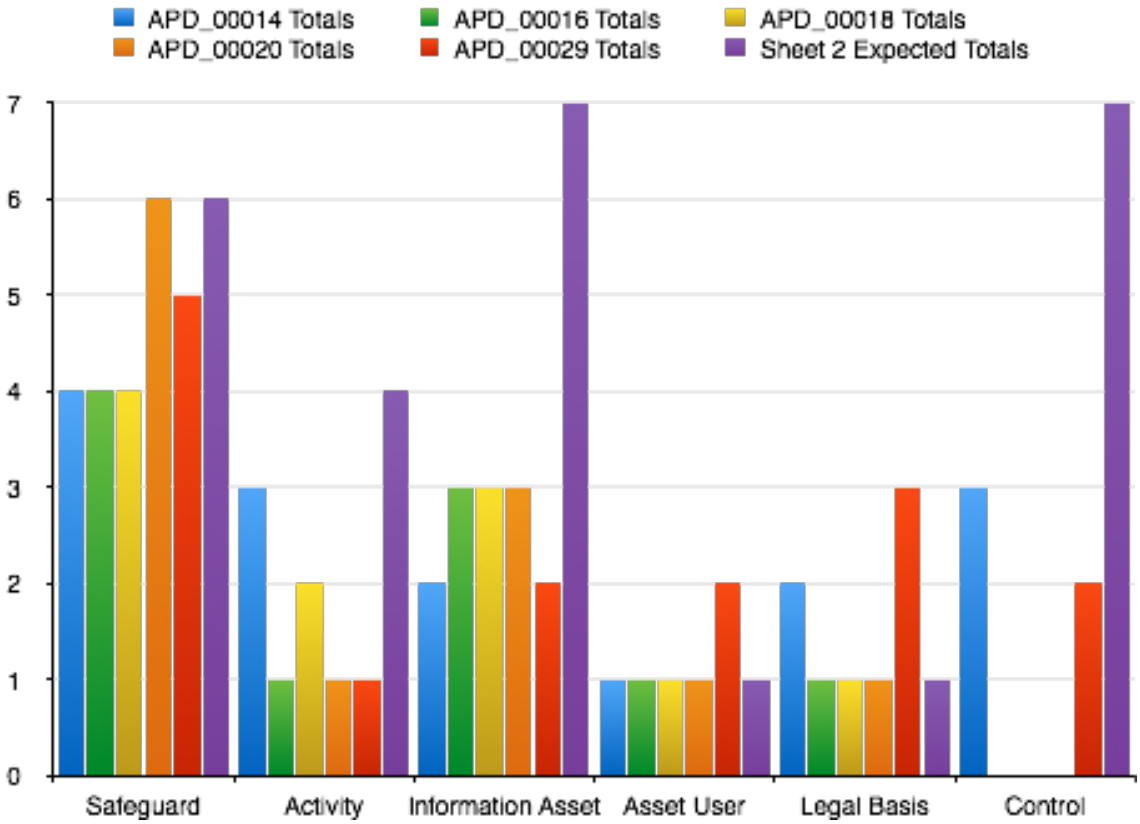
according to the concepts provided by the Secutype model. Graph 5 shows the total numbers of the different concepts added as Compositions for the first experiment by participants who attempted Exercise Sheet 1. This indicates how they interpreted the authoring of policy items using *keibi* and the Secutype concepts and provides a measure of the consistency in interpretation and understanding for authoring the policy.



Graph 5: Numbers of Compositions and Control Clusters Added by All Participants Using Exercise Sheet 1 for All Questions in the First Experiment

There is some variation in the aggregated results across all the Secutype concepts, except Asset Users and Legal Bases, which was expected because the participants were directed to add these directly by the exercise sheet. There is clearly variation in how the participants chose to author the policies using the Secutype models, showing that a consistent interpretation was not supported by the knowledge management approach, which tends to refute the second hypotheses. The greater number of omissions on the part of Participant APD_00021 is again explained by their frustration at losing a sizeable amount of their work in preparing a Safeguard.

The participants also did not meet expectations in the authoring of the policy excerpts, as can be seen by comparing the expected numbers of each Composition, which provides evidence that *keibi* participant behaviour does not encourage expected behaviour when authoring policy items.

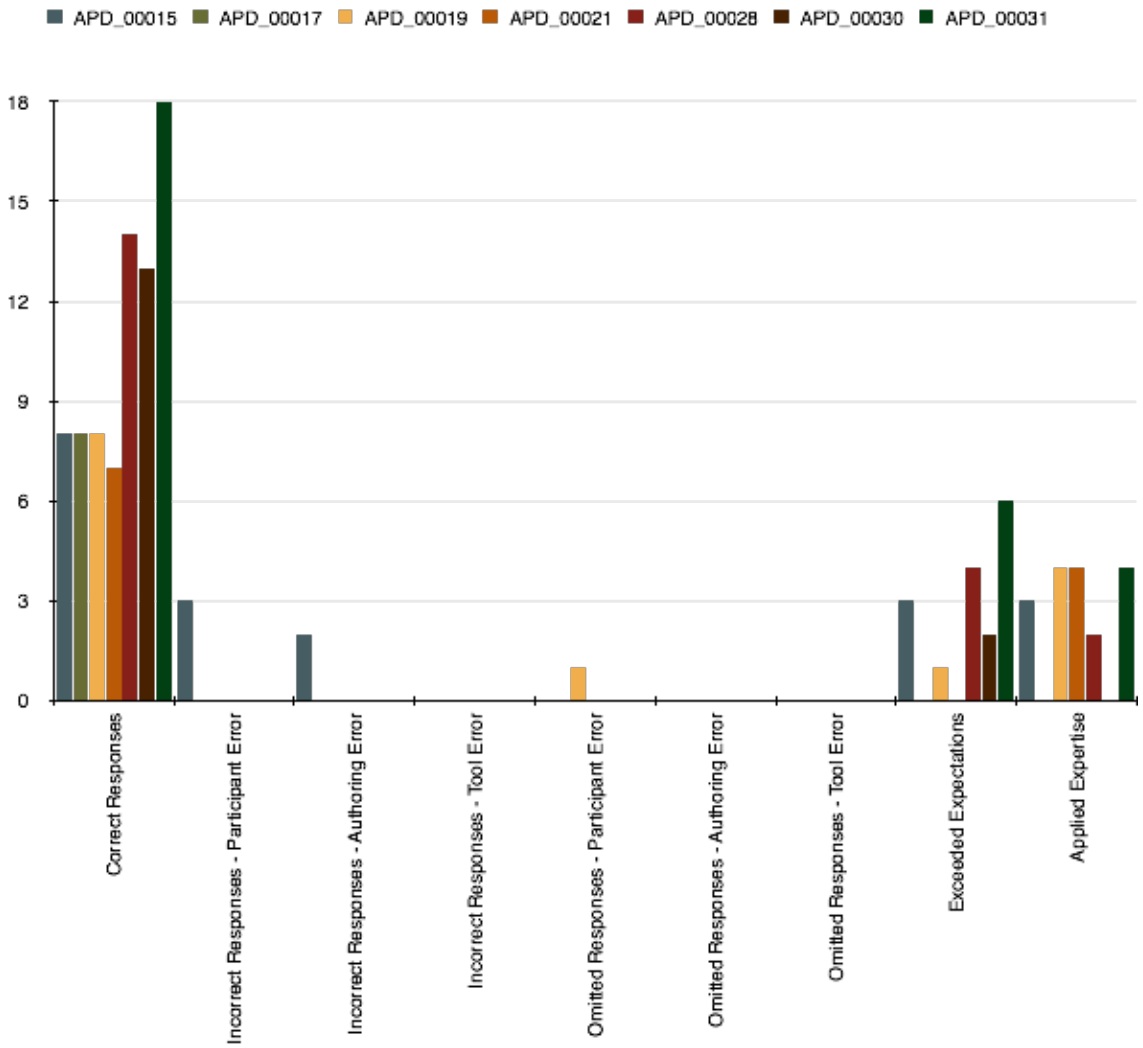


Graph 6: Numbers of Compositions and Control Clusters Added by All Participants Using Exercise Sheet 2 for All Questions in the First Experiment

Graph 6 shows some consistency in the authoring of the policy items, particularly for the Safeguards, Activities and Information Assets. There were again some Control omissions, and a greater use of Asset Users and Legal Bases on the part of APD_00029. This supports the second hypothesis to a greater degree than the results obtained from Exercise Sheet 1, however there are still variations in the responses. The participants who attempted Exercise Sheet 2 also did not respond according to the expectations.

The second experiment measured participant responses to a series of questions about how they would behave with information assets in a set of activities based upon the policies that were authored in *keibi* by their counterparts during the evaluation sessions. The responses were marked against a series of

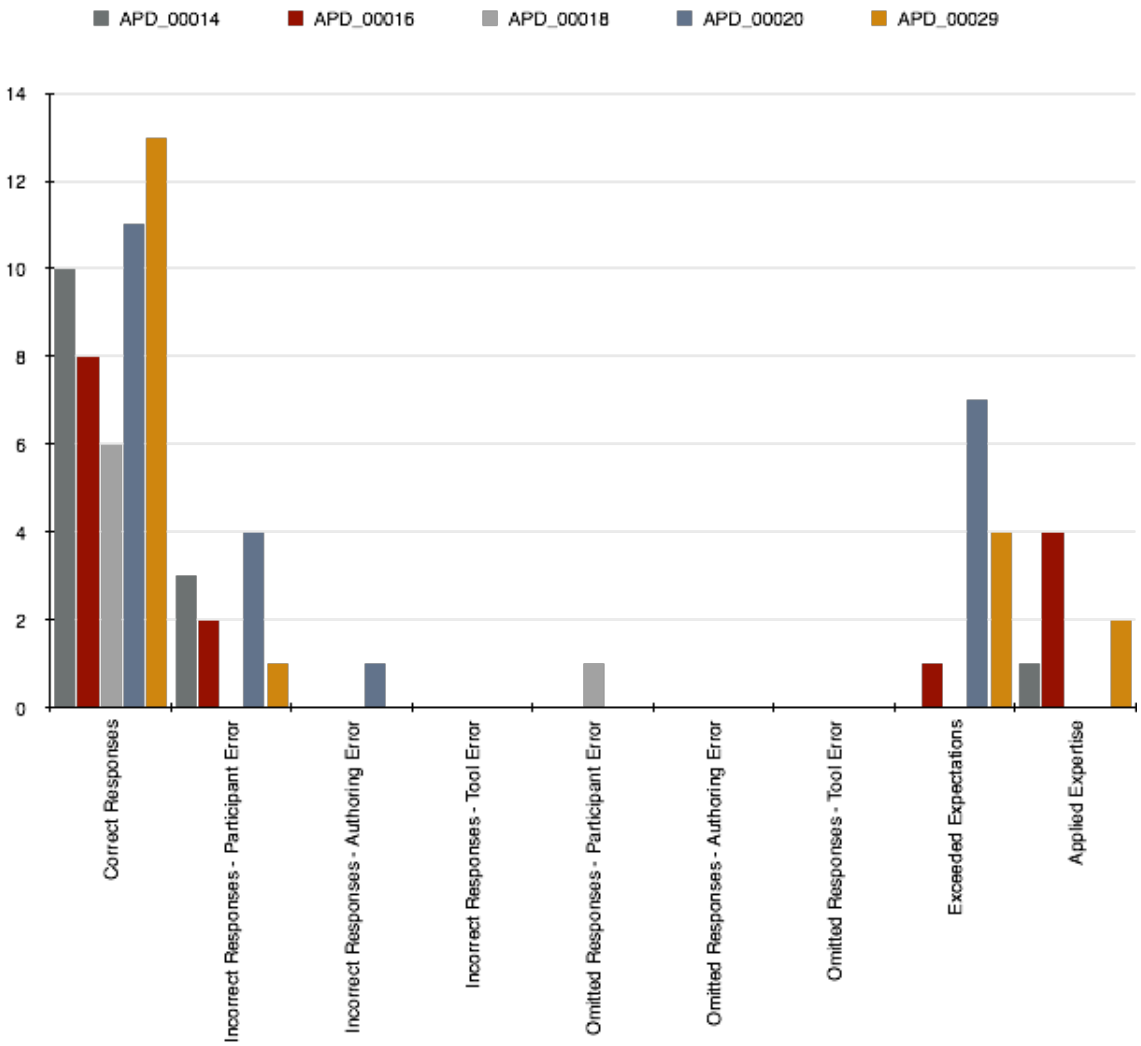
expected answers that had been based upon the expected outcomes specified by participants who had provided the excerpts. Graph 7 and Graph 8 show the aggregated analysis number of correct, incorrect and omitted responses. It also shows the numbers for where participants applied their own expertise and exceeded expectations. A consistent understanding would be demonstrated for reviewing policy by observing a total number of correct, incorrect or omitted examples.



Graph 7: Aggregated Results for Responses During Second Experiment for All Participants Using Exercise Sheet 1 Across all Questions

Graph 7 provides the aggregated results for each of the participants across all questions that were posed in Exercise Sheet 1. This shows a trend for mostly correct responses, a number of which have been placed as a result of participants applying their expertise and exceeding expectations. The number of correct

responses however ranges from seven to eighteen. There is a trend of where four participants APD_00015, APD_00017, APD_00019 and APD_00021 have six or seven correct responses, showing a consistent understanding of the policy items that had been entered by their counterparts. APD_00028 and APD_00030 also had a close number of similar responses, whilst APD_00018 had several more. This seems to suggest a cluster of consistency in interpretation across participants. Across all seven of them there is still variation, however, which tends to refute the second hypothesis that the knowledge management framework limits interpretation of information governance policy, however in this case this did not seem to cause errors and the evidence of participants applying their own expertise and exceeding expectations support hypothesis three, discussed in section 8.6.3. Additionally, the small number of errors or omissions shows that for the reviewing of policy in *keibi* does support understanding of the policy requirements.



Graph 8: Aggregated Results for Responses During Second Experiment Across All Participants Using Exercise Sheet 2 for All Questions

Graph 8 shows the results for Exercise Sheet 2, where there is a similarly low number of errors and omissions, evidence of participants exceeding expectations and applying their own expertise as with the results of the Exercise Sheet 1. There is variation in the numbers of correct answers supplied, which again does not support the second hypothesis. Using *keibi* to review policies does support understanding and use of participant expertise.

8.6.3. Results of Testing Hypothesis 3

The third hypothesis stated that the knowledge model approach supported user expertise when interpreting required behaviour and refined these requirements to computable heuristics. Evidence for testing this hypothesis has been gathered in the analyses performed on the results from Experiments One and Two. These include the results presented in Graph 1 and Graph 2, where it was clear that understanding had been encouraged for the participants and they were able to use the tool with an overall positive effect.

The evidence for expectations being exceeded also supports the hypothesis that participant expertise was supported, and this is further indicated by the results presented in Graph 3 and Graph 4, where the scores for correct responses were much higher than those for incorrect or omitted responses. A similar trend can be seen in the results from the second experiment as shown in Graph 7 and Graph 8: overall there were more correct responses and evidence of expectations being exceeded, along with participants using their own expertise to answer questions. This provides further evidence to support the third hypothesis, where the knowledge management approach helps support user expertise.

The refinement to computable heuristics was not so well supported, however. The fourth question in experiment one was designed to allow for the refinement of the policy items to computable heuristics so that appropriate fields could be removed from a data release in accordance with an existing policy excerpt that had been supplied by the participants. Only one participant, APD_00029, specified the necessary Safeguard with the appropriate Control in a way that could be refined to

a computable policy specification. The response can be seen on page 474. Another participant, APD_00019, also specified a Safeguard with a Control, using a simplified approach to specifying that Control (on page 430). None of the other ten participants specified Safeguards that could be refined to computable heuristics, which is discussed in the next chapter.

8.7. Results from Experiment 3

Experiment three focused on the participant experience, employing a user satisfaction questionnaire and group discussion to gather feedback from participants about their views on the system, its usefulness and their opinions about using it in actual working practice. Participants provided written feedback when they answered the user satisfaction questionnaire, which helped provide further insight into their opinions and experience of using the tool. Some of the items that they raised were explored in more detail during the group discussions. The author noted verbatim the participant responses during the group discussions in each of the evaluation sessions.

This section first presents the results from the user satisfaction questionnaire, providing the scoring applied by each participant to each of the questions as a frequency distribution. This is followed by a discussion of the results for each question and the written feedback participants provided in their responses. The section continues and concludes with a subsection presenting the methods and results of a thematic analysis of the group discussions.

8.7.1. Questionnaire Analysis

The purpose of the user satisfaction questionnaire developed by Lewis (James R. Lewis, 1993) as to apply a validated approach to gathering participant opinions about how they felt about using the system, as well as to prepare them for the group discussion, where their opinions could be explored in more detail. The questionnaire used provided a broad set of questions that would help develop a clearer understanding of participant experience. These included nineteen questions for participants to answer.

This section presents the results from the user satisfaction questionnaire, presenting the results from each question and any written comments that each participant made in the space provided. Lewis developed and validated this questionnaire in line with psychometric factor analyses. The number of participants in this study prohibit a factor analysis: Nunnally recommended that a validation of questionnaires should be attempted with a sample size of participants of at least five times the number of factors (Nunnally, 1978), but since this would require at least ninety-five participants, which is beyond the means and scope of this work, a factor analysis has not been completed. The presentation and analysis of the results instead focus on the scoring provided by participants, and their written comments in response to each of the questions.

The scores for each question were on a seven point Likert scale, where a score of one indicated strong agreement with the statement and a score of seven indicated strong disagreement. Table 30 and Graph 9 below show the frequency of Likert Scale scores provided by each participant for all nineteen questions. These results for each question are then discussed in detail, which in combination with the written comments from participants provided significant insight into the results gathered from the first two experiments. The comments are provided with the discussion below.

Question 1. Overall, I am satisfied with how easy it is to use this system.

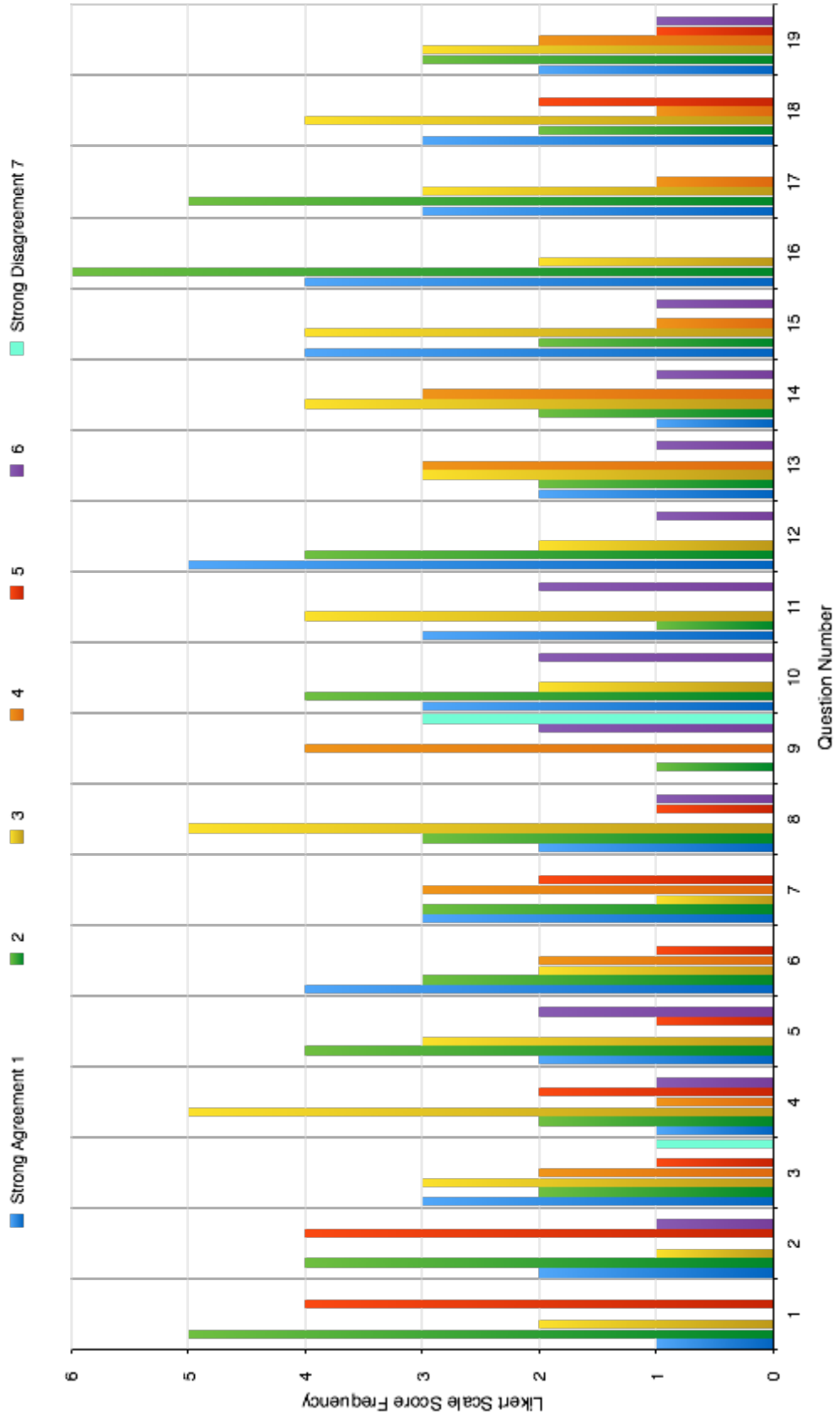
For the first question, the lowest score was one, the highest was five: one participant scored one, five scored two, two scored three and the remaining four scored five. This suggests that participants overall were satisfied with the ease of use of the system. The four participants who disagreed with this statement moderately explained their reasons for this in their written responses below and in the group discussion described in the next section.

Participants made written responses to the first question as follows: APD_00014 felt that "navigation was tricky and needed to switch context (screen) to complete the elements relating to e.g.. Safeguards. The other elements themselves had further elements to add, with their own contexts. Difficult to remember where you are to remember where you are or to remember where to go back to." This would explain their moderate disagreement with the statement.

APD_00017 felt "it's easy to use (I think) although as with most things an example may be useful, having said that the help files are useful." APD_00019 had some problems with the user interface: "apart from a few UI issues, the system is very usable and user friendly." APD_00029 focused on the delivery of the help screens, stating that

Likert Scale Score (1 = Strong Agreement, 7 = Strong Disagreement)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Question							
1. Overall, I am satisfied with how easy it is to use this system.	1	5	2	0	4	0	0
2. It was simple to use this system.	2	4	1	0	4	1	0
3. I could effectively complete the tasks and scenarios using this system.	3	2	3	2	1	0	1
4. I was able to complete the tasks and scenarios quickly using this system.	1	2	5	1	2	1	0
5. I was able to efficiently complete the tasks and scenarios using this system.	2	4	3	0	1	2	0
6. I felt comfortable using this system.	4	3	2	2	1	0	0
7. It was easy to learn to use this system.	3	3	1	3	2	0	0
8. I believe I could become productive quickly using this system.	2	3	5	0	1	1	0
9. The system gave error messages that clearly told me how to fix problems.	0	1	0	4	0	2	3
10. Whenever I made a mistake using the system, I could recover easily and quickly.	3	4	2	0	0	2	0
11. The information (such as on-line help, on-screen messages and other documentation) provided with this system was clear.	3	1	4	0	0	2	0
12. It was easy to find the information I needed.	5	4	2	0	0	1	0
13. The information provided for the system was easy to understand.	2	2	3	3	0	1	0
14. The information was effective in helping me complete the tasks and scenarios.	1	2	4	3	0	1	0
15. The organization of information on the system screens was clear.	4	2	4	1	0	1	0
16. The interface of this system was pleasant.	4	6	2	0	0	0	0
17. I liked using the interface of this system.	3	5	3	1	0	0	0
18. This system has all the functions and capabilities I expect it to have.	3	2	4	1	2	0	0
19. Overall, I am satisfied with this system.	2	3	3	2	1	1	0

Table 30: Frequency Distribution Table Showing Frequency of Likert Scale Scores for Each Questionnaire Question



Graph 9: Frequency Distribution Graph Showing Frequency of Likert Scale Responses for Each Questionnaire Question

"I would add a 'how to in 3 easy steps' section with visual guidance." APD_00031 had some concerns about the interface and interaction, as well as anxiety over whether they knew what they were doing, which might explain their moderate disagreement with the statement. : "When I know what to do, it's OK. Sometimes bit clunky, boxes too small, too many button presses required. But OK. However, I frequently didn't know what to do."

Question 2. It was simple to use this system.

For the second question, there is a similar agreement score applied here, with the same participants disagreeing more strongly as in the first question with one applying a higher score of six, and the same participants agreeing more strongly. The reasons behind this score are revealed below for APD_00014 and APD_00031 where their written comments are presented, and in the group discussions. Seven of the twelve participants agreed that it was simple to use the system. Participant APD_00014 echoed their concerns specified in answering the first question: "as above (response to question 1)/ Pop-up contextual help would have made the system simpler to understand." This might explain their moderate disagreement score. Participant APD_00031: "As above – unclear what I am doing at times. Feels like it would be helpful once I got used to it."

Question 3. I could effectively complete the tasks and scenarios using this system.

The third questionnaire question asked participants to state whether they felt that they could effectively complete the tasks and scenarios using the system. The results show generally greater agreement in this case with more of the participants who felt the system was not so simple to use and were slightly dissatisfied with the use of the system finding that they could effectively complete the tasks. Further written responses from the participants are provided below the graph, which help to expand upon the scores that they gave. Participant APD_00021 strongly disagreed with this, though this is because of their losing an entire Safeguard in the first experiment. Participant APD_00014 made the following comment when responding to the third question: "It wasn't clear what fields were relevant in each case, so was difficult to know if all relevant information had been recorded." This helped to explain their stronger disagreement with the statement. Participant

APD_00015 felt that the statement was "especially true where policy elements already existed. Suspect a little training / practice necessary to build effective policies." APD_00017 expressed concern that they were placing the right details in the correct Compositions: "given (meaning assuming) that I put them in the right place!" APD_00020 said: "there was some disconnect between language used in questions and on the screen. Spent a while looking for the 'Excerpt' button." APD_00031 had concerns about whether they were handling the first experiment correctly: "don't know: Experiment 2, I could do, but Experiment 1, I really didn't know if I was doing the right thing."

Question 4. I was able to complete the tasks and scenarios quickly using this system.

For Question 4 there was a higher frequency of agreement scores but still some stronger disagreement with the statement, where participants APD_00014, APD_00018 and APD_00021 scored quite highly. APD_00014 stated: "as per comments above, a lot of time was spent switching between contexts and trying to remember which elements had been completed / were required," which helps to explain their higher score. The error that APD_00021 made with the Safeguard storage clearly affected their score, rating this as a strong disagreement. APD_00019 felt that "again, some UI enhancements would be beneficial". APD_00015 stated that: "I am a bit tired and this affects my ability to process information logically." There was general agreement that the tool allowed participants to complete their tasks quickly. APD_00031 said "aforementioned comments notwithstanding, reasonably quick."

Question 5. I was able to efficiently complete the tasks and scenarios using this system.

The results for the fifth question, which shows a similar profile to the fourth, has an overall higher proportion of agreement apparent in these results. The disagreement scores are partially explained in the comments made in the questionnaires. APD_00014, who scored higher for disagreement, felt "the queries were easier to complete from the data entry. Though I felt I was stumbling across the answers through the Safeguards section mainly. In the case for Q1 (second section) I couldn't find a suitable reference." APD_00017 echoed their concern

about storing information in the right place "Again I think so; so long as the info was recorded in the correct place." APD_00031 agreed that they were efficient in completing the tasks: "Ditto. (i.e. Aforementioned comments notwithstanding, reasonably quick.)"

Question 6. I felt comfortable using this system.

The frequency distribution of Likert Scale scores for Question 6 shows a general agreement with feeling comfortable using the system, with a similar pattern emerging for those who did not agree with the statement. This is explained by the comments participants made in the questionnaire and group discussions (discussed in the next section). Participant APD_00014 also commented "I thought the system showed potential, though the UI needs to be a lot more helpful, presenting contextual help etc.," whilst APD_00017 highlighted a usual initial discomfort with a new system: "Initially not, but that's the case with any new system." APD_00029 felt that they would be more at ease generally: "more like felt secure in the sense that it is a protocol applying to everyone in the department." APD_00031 expressed their agreement, wondering why they would not feel comfortable: "Sure. Why not?"

Question 7. It was easy to learn to use this system.

The frequency distribution of scores for the seventh question show there was some variation in who disagreed with the statement. Whilst three participants were ambivalent scoring four, seven participants scored in strong or moderate agreement. Participant APD_00015 said "I managed to use it without referring much to the help content. In "real life" I would prepare better by reading through this material." Participants APD_00030 and APD_00031 both did not find the system easy to learn: APD_00031 said in the comments for this question that "I would have liked examples in the help articles." APD_00014 said "I found it difficult to get started, but after the first couple of questions, I started to understand how the sections related to each other."

Question 8. I believe I could become productive quickly using this system.

The frequency distribution for the eighth question showed some further variation in responses from other questions. There was general agreement with the statement, with two cases that disagreed only. Participant APD_00018 and

APD_00029 both disagreed with the statement, where APD_00029 said "the tool does not add any value on my productivity," suggesting that they misunderstood the question. APD_00014 felt that "with more built-in help and mechanism to reduce the need to constantly switch context, it could be an aid to productivity. However, in my particular case, I would not currently have a need for the system (possibly in the future as our service matures)." APD_00030 pointed to some uncertainty about the system use, writing "yes, but I would need more information/guidance. There are many linked sessions and it is not clear how exactly they relate. (Also both NHS Number and CD ROM as Information Asset. Needs some clarification." APD_00031 said that they "(did not) feel like I've appreciated the full utility of the system."

Question 9. The system gave error messages that clearly told me how to fix problems.

The frequency distribution of scores for the ninth question shows less agreement with the statement about the clarity of error messages and potential fixes. Their comments illustrate why this is: APD_00014 said "on 'Activities' the error message 'supplied identifier is not valid for the given property' -> didn't give any context, so process of elimination was needed to establish the cause of the error." APD_00015 said "I made a mistake at one point and the system would not let me proceed. I think I asked Nathan or I guessed that I had to delete "empty" components. The error message did not indicate this." APD_00017 felt "not always – some instances you don't have to click "remove" others you did - not always transparent." APD_00019 said "I only encountered one error message but it did not indicate how to fix the problem or what was the issue" APD_00029: "Got few error messages, and did not clearly tell me how to fix problems. I had to improvise."APD_00030 identified a bug "(Bug)" and APD_00031 "I could broadly guess how to write around them, but they weren't very clear." There is clearly some improvement to be made to the clarity of the error messages, as indicated by these responses.

Question 10. Whenever I made a mistake using the system, I could recover easily and quickly.

For the tenth question there was generally strong agreement with the statement, with the exception of participants APD_00019 and APD_00021. Participant APD_00016 did not feel that they made any errors. APD_00014 said "yes, simple to go back and correct mistakes." APD_00019 reiterated their comment from question 9: "I only encountered one error message but it did not indicate how to fix the problem or what was the issue." APD_00031 claimed that the statement was apt "sometimes…"

Question 11. The information (such as on-line help, on-screen messages and other documentation) provided with this system was clear.

For the eleventh question, there was general agreement here as well, though two participants felt otherwise. APD_00021 disagreed strongly due to the issue with the Safeguard loss. APD_00014 highlighted the helpfulness of contextual help in each Composition screen and some example use cases: "yes - but would have preferred contextual help and some use-cases to illustrate typical scenarios." This was echoed by APD_00019: "The help screens can definitely be expanded with more text and examples - tutorials of each section." APD_00030 echoed the contextual help: "When you are in a screen I prefer a pop-up help text instead of moving back & forth. Also, the help info is not very detailed. E.g. what does the Asset Type Database entail? (Any piece of info or only strictly a database?)."APD_00029 suggested a more visual approach: "The FAQs section could become even more simple. Perhaps consult a designer, add visual material, reduce wordings." APD_00015 was keen to explore this further: "Yes, I look forward to having more time to explore this".

Question 12. It was easy to find the information I needed.

For question 12, there was general agreement, except for Participant APD_00021, who had lost an unsaved Safeguard when they attempted to access the home screen. APD_00014 also added "I found my way around without too many problems, after a slow start."

Question 13. The information provided for the system was easy to understand.

For the thirteenth question, the frequency distribution of scores shows there is some variation in the responses. APD_00020 did not feel that they attempted to access information provided about the system. APD_00021, APD_00030 and

APD_00031 all disagreed with the statement, which seems consistent with other, related questions. Other participants added some points: APD_00014 correctly pointed out that the limited information given prior to starting the experiments was limited: "yes - but limited (deliberately I'm sure!)." APD_00015 said "I didn't explore this. I look forward to doing so." APD_00018 echoed the need for examples to help make system use clearer: "examples in the text explaining the different sections would have helped." APD_00028 felt that the tool "could do with a short tutorial at first to explain the concepts, but the help section is very good." APD_00029 suggested some targeted training: "this has to do with the person that inputs the info, so it is user driven mainly. So I would suggest that this person has some training on how to properly use it instead of just reading the FAQs by himself."

Question 14. The information was effective in helping me complete the tasks and scenarios.

For the fourteenth question the frequency distribution of scores shows there was again a general agreement with the statement, and disagreement from the same participants as the previous question. Participant APD_00020 did not access any information for the system again. APD_00014 added that the information for the system was limited: "as above (i.e. Yes - but limited (deliberately I'm sure!))." APD_00015 also reiterated their comment from the previous question: "as above (i.e. I didn't explore this. I look forward to doing so)." APD_00019 felt "help screens can be expanded." This indicates that participants felt that some of the help and system information could be expanded, consistently with responses made in previous questions.

Question 15. The organization of information on the system screens was clear.

For Question 15 the frequency distribution of scores again shows a general agreement with the statement, with some disagreement from participants APD_00021 and APD_00031. There were some supportive comments: APD_00014 said "Yes," APD_00015 "Yes, very good" and APD_00019: "Good layout - would benefit from tooltips." This indicated that participants generally felt that information on the screens in general was well laid out.

Question 16. The interface of this system was pleasant.

The frequency distribution for the sixteenth question shows a strong agreement amongst participants that the system interface was pleasant. Additional comments added some further details. APD_00015 said "I noticed when I was reviewing a record, sometimes a bow on the right floated in temporarily labelled view record. Was this intentional?" The participant is referring to a system feature to make clear that the icon allows you to view the record. APD_00018 said "The free text boxes are small, so difficult to read long excerpts." APD_00019 offered these compliments: "Good colours and layout." APD_00029 suggested improvements "In the future the user interface could get better. I would suggest the use of a designer in order to redesign all the screens. And I could not find how to zoom in/out. Settings section should be added." Generally participants were satisfied with the interface, but improvements could be made.

Question 17. I liked using the interface of this system.

The frequency distribution of scores for the seventeenth question again shows an overall strong agreement, with the exception of participant APD_00018. APD_00014 also commented "apart from context switching detailed elsewhere" and APD_00029 "the small letters in every wording did not help me enjoying it."

Question 18. This system has all the functions and capabilities I expect it to have.

This is also true for the statement in the eighteenth question, with the exception of participants APD_00014, who referred back to their comments elsewhere in the responses, and APD_00031 who said: "Don't really know what I do expect!" Other comments included from APD_00015 "would like to use again," APD_00018: "maybe have another section on data transfer?" (which itself is an Activity), APD_00019 "would be great to have a "generate policy " button that would group all information in a single document for sharing" and APD_00028: "Multiple Asset Users per Activity?" (which is of course possible).

Question 19. Overall, I am satisfied with this system.

The frequency distribution of scores for the nineteenth question shows that eight of the participants felt that they were satisfied overall, two were ambivalent and two disagreed, with APD_00021 disagreeing quite strongly due to their frustrating experience. Additional comments were as follows:

APD_00014: “With further development, yes.”

APD_00015: “Yes, would like to use again.”

APD_00018: “I am not sure about the advantage compared with existing policy documents, where all excerpts and information assets are already included.”

APD_00019: “Great work :-).”

APD_00030: “I think that 1, well-worked out example, would elucidate the function of each section, and how they all relate to one another.”

APD_00031: “Need more training and fewer bugs! ☺”

Many of these comments reiterate what was said in the comments for the previous questions.

8.7.2. Thematic Analysis of Group Discussions

Each evaluation session completed with a group discussion to allow participants to further reflect on the use of *keibi*. The discussion was semi structured, where four general questions were used to frame the discussion and permit further exploration of their views around their experience of using the tool. This section provides an overview of the approach used to conduct the thematic analysis and discusses the themes that were identified from the group discussions. This includes a description of how data was collected and transcribed, analysed to identify repeating themes in the data and generate codes, development of, review and definition of themes based on how these codes interrelated, how the emerging themes illustrate the effectiveness of the knowledge management approach and tool as a proof of concept solution and the analysis of the data based upon the refined themes after this process. The section concludes with a discussion of the themes that represent how the participants viewed the use of the knowledge management approach and their thoughts about its effectiveness to validate it as a proof of concept.

Data was collected during the group discussions by noting verbatim the responses of each participant to each question, where the notes included the participant identifier along with how they answered the question. There were only two or three participants present for each session and the discussions sessions did not last longer than thirty-five minutes, making this approach to data

collection reliable. The noted responses were repeated back to the participants after they had answered the question where clarification was needed if they had not been heard, their answer was unclear or they were not fully understood. Follow up questions were noted when posed, and the responses were noted as described above. Questions posed by the participants and responses to them were also noted. At the end of each session, the notes were written up in a narrative form, which can be found under Appendix 20.

The notes were read and re-read both during the transcription process and after it had been completed to identify recurring themes in the data, which was achieved by coding the data sets. This involved identifying relevant, meaningful items relating to the use of the tool and the knowledge management approach's effect on participants, whereupon the data could be categorised and used to help develop themes. An initial set of codes was identified during the transcription process and first readings, after which the software tool NVivo was used to code data items and further refine the categorisation where appropriate so that relationships could be established between the coded data and themes could start to be distilled.

To illustrate, the code "Learning Curve" was identified as an appropriate categorisation based upon the responses from several participants, for example APD_00018's response that they found the system hard to use initially because they could not understand what the concepts meant. This was also referred to by APD_00015's. These in turn were consistent with APD_00014's statement that they experienced a "steep learning curve to begin with" in understanding how the concepts related to each other, which APD_00019 agreed with. This was clear also from the response by participants APD_00016, APD_00017 and APD_00021, who felt that their initial interactions were trial and error because they were not sure about the meaning of the concepts initially, leading to an initial coding of "Trial and Error" that could then be subsumed into "Learning Curve." Participants APD_00028 and APD_00031 also felt that it was initially difficult to identify how to "encode" the concepts in the tool. APD_00030 claimed that "was pretty easy to learn, but you need to learn some elements that are not intuitive, for instance they were not clear on what details to put in the different concept screens"

This led to the definition of "Learning Curve" as a code. Additionally, "Understanding Concepts," "Concept Relationships" and "Understanding the Tool" were also identified as appropriate codes. The relationships between these and "Learning Curve" helped to identify themes in the data: participants initially found it hard to understand the concepts and their relationships between each other, with some finding it unintuitive, and this led to another recurring theme where participants found there was a steep learning curve for using the knowledge management approach initially and when using the tool. Another theme that emerged was that participants found the tool easier to use after they had some practice with it, or as participants APD_0028 and APD_00031 put it "as they went through the questions, they began to understand how the various parts went together." These led to the conclusions that using the knowledge management approach had a steep learning curve because of an initial uncertainty over the relationships between the different concepts and once overcome, participants found the tool easy to use after some practice and developing an understanding of how the concepts fitted together. This further went on to support another theme that emerged, where they found it thought provoking in the area of information governance.

This process continued by reviewing and re-analysing the transcripts using NVivo, further developing the coding and identifying the relationships so that themes could be developed. As themes were being developed, they were refined so that they could be used to provide meaningful statements about how the approach affected participants when they applied themselves to authoring and responding to information governance policies. This produced eight other themes, which are discussed and explained below by reference to the coding of the source data in the transcriptions to explaining the results of evaluating the hypotheses and validating the tool as a proof of concept. The data has been categorised and the themes have been expressed in terms that were provided by the participants during their responses during the group discussions.

Another theme that emerged related to the learning curve difficulty. This was identified as navigation between compositions and also accessing help screens being frustrating. Both participants APD_00015 and APD_00019 felt that the

"biggest thing was that you had to switch between contexts and even had to drop out of what you were doing to access help screens" whilst APD_0017 and APD_0021 "felt that coming out of the editing screens to view the help screens was a nuisance - they wanted to crack on with the exercise." As discussed in previous sections, APD_00021 "got caught out" when they navigated away from the Safeguard editing screen to consult the help screens. APD_00031 felt navigating away from the editing of concepts was unintuitive and APD_00028 felt the interface "did not lead you through in an expected order." Participants also agreed that in addition to navigating away from editing to visit help screens, having to leave the editing of a Safeguard to add a new Information Asset, for example, interrupted the flow of the activity they were engaged with.

This theme related to another reoccurring theme amongst the participants, which was their noting a heavy workload when authoring the policy items in *keibi* to begin with. Data items were coded as "Workload," "New Project Setup" and "Granularity of Detail" which helped to identify this theme. Participants felt that this approach would be particularly useful for new projects that were starting to define their security policies, whereas the work would be replicated in cases where existing policy was already present, as mentioned by participant APD_00015 and APD_00016. Statements from APD_00018 and APD_00028 who felt that entering existing documentation "would be a huge duplication of effort" also supported this. APD_00029 and APD_00030 felt that the tool "involved a lot of typing" and the workload could be reduced with greater use of selection and tick boxes.

Responses from Participants APD_00014 and APD_00019 also picked up on this theme, where they felt that the level of detail offered by the knowledge model was "too granular," which is consistent with APD_00021 and APD_00031's responding that "it was not clear how practicable in a large scale framework" authoring to such a level of detail would be. APD_00031 also felt that they "did not feel that all the fields were necessary and they found this a bit confusing." which could in and of itself cause confusion and added to the workload. They and others also recognised that the level of detail made people think about information governance issues: APD_00031 and APD_00020 both recognised this and felt that

ISO 27001 made organisation members think in that level of detail. APD_00030 said that provided someone else entered the details, they felt that the tool would be very useful in guiding their behaviour.

Recognising the heavy authoring workload also had important benefits showed a relationship with another theme, where participants stated that interactive help screens, wizards and tool tips to guide policy development would be useful. This theme emerged from coding "Contextual Help," "Guidance" and "Wizard." The need and potential to further assist the development of policy using these kinds of help elements was a clearly reoccurring theme. Participants APD_00016, APD_00019, APD_00020 and APD_00028 said that wizards would be very helpful in guiding policy authoring and developing details as required by Data Management Plans. APD_00030 and APD_00029 both felt that whilst the existing help screens were "pretty clear," it could reduce the number of words and perhaps employ visual materials to aid better guidance, adding that a designer could be used to achieve this. They added that a "tiny question mark" could be used to bring up a popup help screen for each field in the editing screens to offer this guidance. APD_00020, APD_00028 and APD_00031 all felt that flow diagrams would be helpful in guiding users on their expected behaviour. APD_00028 referred to the possibility of different users having their own labelling conventions without guidance. Participants felt that this would make an effective advisory system with some additional guidance mechanisms as described.

The helpfulness and usefulness of the tool were established in another theme that emerged from the responses. This theme was the usefulness of having a consistent view of the tool and how it encouraged collaboration between users, which would help to improve current practice. The coding of data to "Collaboration," "Completeness," "Consistent View," "Guidance" and "Document Management" combined with their interrelationships helped to develop this theme. Participants APD_00020, APD_00028 and APD_00031 all felt that this would be an important resource for Principal Investigators and would form the basis for collaborative working with different users as well as a document management system. This was echoed by claiming that this tool offered a means to "advise collaborators on how they should behave and what they needed to know"

when working with shared healthcare data. APD_00016, APD_00017 and APD_00021 all felt that this tool would help ensure that staff "sung from the same hymn sheet" in terms of offering a consistent view of what was expected of them, whilst APD_00019 also stated that "having other users know about policy items rather than just managers would be helpful."

A theme around issues with current practice emerged as participants discussed the usefulness of the tool and how they thought it could help. Issues relating to showing good practice, understanding engagement and adherence were amongst some examples raised across the participants. They felt that the approach offered a means to provide reassurance that good practice was being adhered to. They identified it as being able to handle risk assessment, and felt it would offer a much more pleasing interface than an existing risk assessment tool. They also felt that the tool would save people from having to go through "pages of documentation," which they felt did not always happen as expected, and would help them to make fewer mistakes. They identified that the approach offered an alternative to having to have expected behaviour "burned on to [researchers'] brains" to be applied and in practice as they saw fit, with no easily accessible means to check policies or show good working practice, where they felt that people did not actually read the documentation.

This theme also related to two others, in particular most of the participants' recognising that the approach would allow for a means to tailor / personalise and summarise information governance requirements and policy details to individual users, which would greatly help them understand their responsibilities and what was expected from them, by reducing the amount of information that they had to sift through. Participant APD_00031 felt that "you should only see what you need", whilst APD_00014 said that it would be very helpful to have a policy generation tool, which could summarise Safeguards and put all the important information together. APD_00017, APD_00018 and APD_00021 agreed that the knowledge modelling approach would also help to manage the issues that emerged in themes around aiding users in deciding how best to manage the expectations and trade off between pragmatism and governance needs.

This related to another theme that emerged, where participants identified the auditing features as being particularly useful for demonstrating compliance and adherence with policy. Participant APD_00016 for instance pointed out that this would "capture elements of governance and security for external review" and could also flag examples of non-compliance and areas for improvement. APD_00018 felt that the auditing features could provide proof that users had actually looked at policy items. Participants APD_00020 also saw that this ability to show compliance and record policy stipulations would be particularly helpful in demonstrating compliance with the Information Governance Toolkit and ISO 27001 certification, as well as developing Data Management Plans. APD_00016, APD_00017 and APD_00021 also felt that the tool could provide reassurance that good practice was being adhered to.

One theme emerged from responses by participants APD_00015, APD_00018, APD_00020, APD_00021, APD_00029, APD_00030 and APD_00031 who thought that the approach and tool would be particularly useful for training and education. Participants elaborated on this and felt that this tool would be important for the induction of new staff, where the consistent view offered by the approach and tool of what was expected would help to educate and train users, particularly for induction of new staff members into a project. This helped to validate the tool as a proof of concept, where the participants identifying it as a potentially very useful tool for education showed that it would be applicable in working practice, as well as help to encourage understanding and good practice when used for guiding researchers on how to work with healthcare information.

The themes that have been identified from the group discussions point to some issues regarding the navigation using the tool and an initial uncertainty over the meaning of the concepts and their relationship with one another. These were nevertheless overcome after some practice with the tool and overall participants were very positive in their feedback. They generally felt that the look and feel of the tool was pleasing, however they recognised that the approach and implementation of the tool would provide guidance that would help develop a consistent approach for authoring policy, subject to some additional interactive help screens and wizards, would provide a consistent view of the policy items that

had been authored in it, and that it was thought provoking, encouraging users to think about issues around information governance based on the level of detail that they would have to go into in order to author and use it effectively.

Participants also identified many of the issues that this thesis has reviewed, recognising that the tool would be helpful in handling them. This included a means to summarise and simplify the representation of the policy items, as well as reduce the amount of reading users had to do, whilst also providing a means to check whether users had actually read to required details. With further work, they identified that the tool could be tailored for individual users, to further promote understanding of what was required of them. Participants also recognised that the tool would be particularly useful in guiding collaborators on how to behave with shared information, offering a shared, consistent view across collaborating users. They felt that the tool would help provide reassurance that people were behaving as they were expected and guided effectively.

These themes that emerged from the group discussion help to support the hypotheses as well as validate the approach and tool as a proof of concept. Participants felt supported and able to use the tool, understanding it and the knowledge management approach as they gained further practice. It is clear that further development on the tool itself will be beneficial, particularly in the area of direct guidance using interactive help mechanisms for authoring and explaining the concepts that form part of information governance as a means to offer advisory and decision support.

8.8. Summary of Hypotheses and Proof of Concept Evaluations

This chapter has provided the approach, methods and analysis of the results of the evaluations of the tool. The first two experiments generated results that were necessary to evaluate the hypotheses and establish the evidence for supporting the tool usefulness as a proof of concept. The third experiment completed the evidence basis for evaluating the tool as a proof of concept by focusing on the participant experience and their opinions of using the tool. The results have been analysed and used to test the hypotheses that were proposed in addition to validating the tool as a proof of concept solution. This section summarises the

analyses and provides the conclusions that can be drawn from each of the evaluations.

8.8.1. First Hypothesis Summary

The first experiment tested the first hypothesis, focusing on whether the tool encouraged understanding as well as consistency in that understanding across the participants. The second experiment tested this hypothesis to see whether policy items authored by the participants helped their colleagues understand how to handle situations that involved handling sensitive information, which provided an additional opportunity to see whether a consistent understanding could be achieved using the specified policy excerpts. The third experiment allowed further results to be gathered based upon the participants' own experience, which would be important to provide further explanation of the results gathered in the first two experiments.

The results from the analysis clearly show that understanding was encouraged across the participants across both experiments one and two, though this was not always consistent and some participants demonstrated more examples of understanding than others. Whilst there was some evidence of misunderstanding, further analysis of the results showed that these were related to minor errors that were not critical to the overall expression or subsequent use of the policy items that had been managed in *keibi*, particularly since there was an overall indication that participants understood what *keibi* presented them and how they should use the details as shown in the second experiment.

Further exploration through the user satisfaction questionnaires and groups discussions showed that the participants felt that they needed more guidance on how to specify policies using *keibi* and that a wizard and interactive help screens would help this. Participants were also not fully confident that they understood the relationship between what was in the policy excerpts and how to use the Secutype Classes to express them. This helped to explain why there was inconsistency in understanding when authoring the policy items. The second experiment also showed that *keibi* encouraged the understanding of reviewing and

using policy items, but there was an inconsistency in the levels of that understanding.

The first hypothesis was partially supported in that it was clear that *keibi* promoted understanding but was partially refuted in that the understanding was not consistent across the participants. Ensuring that a more consistent understanding across users can be achieved relies on the availability guidance and interactive help screens, which can help to achieve a more consistent understanding according to a predetermined set of criteria for defining the desired understanding and consistency across users.

8.8.2. Second and Third Hypothesis Summary

The first and second experiments primarily tested the second hypothesis, focussing on whether the tool limited variation in both authoring policy items and interpreting them. This was established by analysing how the participants used the Secutype knowledge model to express the policy items in *keibi* in the first experiment and how their counterparts used the authored excerpts to answer the questions posed in the second question.

It was clear from the results that variation was evident in both authoring policy and using that policy as advisory on handling sensitive information. This leads to the conclusion that the second hypothesis has been refuted. It is clear that a knowledge management solution does not in and of itself limit variation in interpretation. Based on the responses to the user satisfaction questionnaire and group discussions, the knowledge management approach does provide a basis for helping to develop a common interpretation of information governance requirements and encourages user understanding of these. The support for using the tool as an educational and new staff induction resource supports this basis, and the development of more interactive guidance screens can assist with developing the common understanding that existing solutions currently lack.

The third hypothesis stated that the knowledge management approach would support participant expertise for meeting the expected requirements of information security governance so that these could be refined to computable heuristics. Both the first and second experiments provided the focus for evaluating

the third hypothesis. It was clear that *keibi* supported participant expertise, as shown by the evidence gathered for participants exceeding expectations and applying their own expertise.

It was however clear that it did not support participants in specifying the policy requirements in a manner that would allow them to be computable for software configuration in the majority of cases. In order to achieve this, it is clear that further guidance is needed to help participants specify policies in this manner. Whilst the details are clearly held in a computed state, further refinement would be needed in order to use them to configure the software tools such as those responsible for de-identification or access control. Though the example in section 7.3 provide evidence of the computability of the Secutype models for access control and privilege management and user expertise is supported by the tool, the third hypothesis has been partially validated, because *keibi* did not guide the participants to specify policy in a way that would support refinement to computable heuristics.

Chapter 9. Discussion

The research work in this thesis included several areas of inquiry. They pertained to a literature review and a review of legislation, guidelines and standards for processing sensitive information described in Chapter 3. Chapter 4 focused on the information strategy to support the sharing of electronic healthcare record information for care provision and clinical research. A more practical focus was possible through a series of case studies, which involved clinical research projects that required the development of clinical record databases and the development of electronic healthcare record systems for use in clinical practice and disease registry development, described in Chapter 5. These helped to gather and specify requirements for the proposed knowledge management solution, which were analysed to design, develop, implement and test that solution, as described in Chapters 6 and 7. An evaluation using live participants of the resulting Secutype knowledge model and *keibi* tool tested the hypotheses proposed by this work and the effectiveness of the tool as a proof of concept solution for managing information governance requirements as described in Chapter 8. This chapter discusses this research work by considering the learning outcomes from and challenges of the literature and information governance requirements reviews, the case studies, the knowledge management solution design, development and implementation process and the evaluation of the tool in testing the hypothesis and validating the thesis. It concludes with a discussion of the strengths and limitations of the work.

9.1. Literature and Information Governance Requirements Sources Review

The research work reported in this thesis has reviewed the literature from several disciplines. These have included those relating to the broad societal expectations for protecting healthcare information, which focus on the challenges to legal, good practice and ethical bases for processing electronic healthcare information and reusing it for research. The area of information security and technical approaches has also been explored, where there has been a clear focus on protecting the

identity of participants in research projects and anxieties over the level of identity protection assurance that these approaches actually provide. In addition to these approaches, current methods for protecting information at the level of human understanding and interaction has focused on the development of policies and guidelines, where discussion about the effectiveness of these policies, the concerns about a common understanding of the contents and the effectiveness of these approaches have been considered.

Public and professional anxieties over the sharing of sensitive healthcare records have also been explored as part of the literature review, particularly in light of the more recent examples of data sharing anxieties. The review has therefore also included the literature pertaining to the increased support for clinical research in terms of funding and infrastructure. This led to a review of the literature describing development and engineering of interoperable electronic healthcare records for the purposes of sharing that information with a consistent understanding and meaning across the different users of that information, which have helped to support the reuse of these records for purposes other than clinical care, including research.

The learning outcomes from the literature included a clear uncertainty over the effectiveness of existing protection measures and the constant need for update and systems development that should be in place. Whilst various techniques of de-identification and encryption have been largely shown to be fallible and offer less assurance about the protection measures that have been proposed, the literature also shows a large amount of consideration regarding the use of access control, including roles based, and privilege management. These approaches have been identified as being unable to cope with the requirements of programmes such as Connecting for Health as was, and whilst research continues into these areas, they do not tackle the core issue that actual working practice is not being represented by the simplified access control models, that access is being denied unnecessarily, or that it can and sometimes should be overridden.

The risks involved with sharing information are clear from the literature: identities cannot be assuredly protected, access control and privilege management is complex but still too basic, or simplified yet implemented in a complex fashion.

Encryption approaches can be broken and people make mistakes and lose data. The support for sharing has nevertheless continued and increased, particularly in terms of funding and changes to legislation, which are decried by privacy advocates and clinical professionals alike as contravening data protection and human rights legislation.

There are however gaps in the literature for this area. There is a lack of published work on developing a unified understanding of and integrating the core components of information governance, including the legal foundations for information processing and protecting individuals through to the ethical considerations and development of sound research and finally the enactment and control of good working practice using information security techniques and management. The challenges of ensuring legal compliance and effective control have not been reconciled with the clear fallibility of the technical information security mechanisms, beyond a skew towards offering a host of de-identification strategies that have had the effectiveness of the anonymity they offer repeatedly questioned. There seems to be no clear answer regarding the legal compliance of various technical security measures much less a consistent and combined appreciation of the relationships between the legal basis for processing sensitive information, the ethical requirements for creditable clinical research, the enactment of appropriate information security policy and use of technical security measures to protect information and the people about whom it has been captured.

The literature does not acknowledge the risks that remain to participants and the need identified by the author and via conference attendance about engaging directly with people who are processing information with a view to educating them about good practice and information governance as opposed to simply training them. It also does not provide a solid basis for engaging with the public about the risks that are involved but also the potential benefits of reusing healthcare records for research. Much of the discussion and debate is being handled within the media, but it is clearly difficult for a scientifically valid pursuit of work to continue where there is a lack of common agreement about the bases for protecting EHRs when used in research, inconsistent definitions of information governance and a denial about the risks involved with processing these records for

use within the research and other contexts, which remain despite the use of protection measures.

There was limited discussion in the literature surrounding the effects on people when using existing information governance frameworks and facilities. There is a more general acknowledgement that the existing process is not ideal, however a means of handling the issues has been left to the informatics and clinical research communities to consider possible solutions. Many of the legislative changes, procedural updates and proposed solutions have remained largely reactive to sharing agendas or public and professional anxieties, where there has been little evidence to date that the proposals, debates and implementations have poorly understood the requirements and the need to offer an understandable and consistent approach to managing information governance requirements. There has however been a richer set of literature that emphasises the importance of public engagement, transparency and demonstration that good working practice has been achieved in handling electronic healthcare records when used for research.

The literature and information governance requirements sources have made clear that the sharing of information to support care and research is a high priority to the UK and European governments. This was clearly seen in the publications about research uses to date, as well as the sources describing the information strategies for the NHS. There was also evidence that over anxious information governance approaches were hindering research and other sharing, as shown in the second Caldicott review. This pointed to a lack of understanding about information governance requirements and expectations, and a lack of engagement in the areas of both sharing provision and the reasons as to why information processors were apprehensive about sharing information.

It was clear that the work performed in the area of developing knowledge management services for managing healthcare information provided a powerful solution to help provide a consistent view of healthcare information and support the sharing of that information for care and wider purposes. The integration of record structure, facilities to develop consistent understanding using constraint models and semantic services for applying international coding schemes where

they exist have been pivotal in realising establishment of a proven knowledge management approach that supports healthcare provision and the associated secondary uses. These engineering approaches are however complex and there is a significant difficulty in not only understanding the modelling approaches, but also the development of these systems. There is also very literature pertaining to the effectiveness of this approach on that care provision, as well as the utility for clinical professionals who provide care and the patients who receive it.

9.2. Case Studies

The case studies featured examples of clinical research projects that were dependent on the use of EHRs or information derived from those records to answer a variety of clinical research questions, support healthcare policy, assist in commissioning of services or clinical trials across a variety of chronic and acute conditions and treatment services. They also provided an insight into the development and implementation of EHR systems for clinical care and disease registries, which managed EHRs for purposes that were focused on clinical care or the development of registries that would be used to support clinical trials recruitment or other clinical research projects. This allowed the author to explore the area by approaching and discussing information governance concerns with a wide variety of stakeholders, including those who were pivotal to running healthcare information systems, supporting research repositories and conducting research. There were several patient and public engagement opportunities where the patient perspective could be investigated directly through engagement and interaction with treatment advocate groups and advisory boards. The case studies also provided a means to investigate UK as well as European perspectives, where inferences could be drawn about consistencies across European Union member states.

The case studies provided practical experience of using the established clinical knowledge management approaches in practice for developing research and clinical care systems. This allowed a number of key areas of insight into their development and deployment in practice. Whilst powerful and useful for practice, their development was nevertheless complex in terms of understanding the

modelling approach as well as developing systems that could be used to treat patients and support research. This emphasised the need to simplify the process of developing knowledge models and implementing systems for not only clinical care, but also the proposed approaching the context of information governance for clinical research. It was also possible to see the process for developing and designing knowledge models across a series of domain experts, where the process was identified as lengthy and complex itself. This helped to provide a frame of reference for defining the process by which the Secutype model could be developed and used.

The case studies also provided practical experience in managing information governance requirements for clinical research. The key insights gained from the case studies included a clear lack of support and understanding of the combined legal, ethical and practical needs for people conducting research, a haphazard process for gathering appropriate permissions and incorporating those into working practice. This also illustrated varying appreciation of the responsibilities that applied to each person who handled the information, and their own anxieties about not having a clear, concise and meaningful reference about how they were supposed to behave and what their responsibilities were.

The case studies helped to provide insight into the process of applying the core principles of the ISO 27000 series of standards for information security, including experience in running risk assessments and developing the recommended information security management systems. This provided the author first hand experience in the process of refining the results of the risk assessments and gathering the required knowledge artefacts to develop policy and apply that in practice. The key insights that were provided by this experience showed areas where a knowledge management approach would in theory be particularly useful, by automating and simplifying the process of developing policy and providing a means for users of that policy to have a single point of reference for guiding them on how to behave with sensitive information.

9.3. Information Governance Knowledge Management Solution Requirements

By reviewing a series of legislative instruments, good practice guidelines and international standards in combination with the stakeholder interaction during the case studies, it was possible to develop a series of requirements to develop a knowledge management system for use in the context of information governance. These sources provided an insight into the kinds of information that needed to be captured and how it should be structured, organised and presented to inform the knowledge that would effectively support policy writers and users in how to develop and use the policy elements that the system would store. The sources were supported by further insights gathered from the case studies, which helped the author to place the gathered requirements into the context of developing the knowledge model as well as the wider framework that would be used to provide the policy advisory tool.

The requirements have been categorised according to the process of developing knowledge models and using them within a knowledge management framework so that appropriate record keeping and advisory tooling can be developed. This has involved developing the requirements for a constraint model that met the needs of the information governance domain, as well as those to provide a means to develop that model and use it to generate the appropriate tool using a knowledge management framework. The use of ontologies within the area of knowledge management is an important approach for developing semantic interoperability and helping to develop a consistent understanding of healthcare concepts. There is however a lack of consensus on terms and concepts presented in a coded form, for the area of information governance.

Whilst it has therefore not yet been feasible to develop an ontological relationship of the concepts, the knowledge management approach that has been developed provides a basis and opportunity to develop such an ontology. This forms the basis for potential further work, where the development of an agreed terminology and the incorporation of appropriate ontologies to develop the relationships between these concepts in line with the knowledge models could

help to establish a tighter relationship between the knowledge modelled concepts of the Safeguard, Activity, Information Asset, Legal Basis and Asset User. This is also potentially desirable for helping to organise, quality assure the semantics and cross reference a growing library of Secutype concepts as well as assist with the guidance that the knowledge management approach provides, which represents a further area of requirements development discussed in section 10.1.1 as potential further work.

Whilst the background engineering principles for developing knowledge management are complex across the various standards that exist, the methodology and approach represented a robust method for capturing and sharing information consistently and reliably for the variety of users who need to remain informed about information governance. Focusing on the knowledge model and wider framework requirements meant that the author was able to adapt existing, proven elements of the knowledge management facilities in the clinical context to the information governance domain, simplifying the development, implementation and process of adapting knowledge models into use by live participants. By applying an iterative approach to the development, testing and evaluation, much has been learned about the effectiveness of the knowledge model and the requirements upon which its design has been based. Whilst the legislation, expectations and guidelines will continue to change, the requirements that have been gathered allow for these variations and any fundamental amendments can be applied through the Unified Software Development Process (USDP). This also allows for further development and refinement where applicable, and potential areas are discussed in the next chapter.

9.4. Design, Development and Implementation of the Knowledge Management Solution

The development process followed the established USDP, providing the author the means to assess the background literature, materials and learning from the case studies, and organise the learning outcomes from these into a software development process, starting with the requirements gathering and development discussed in the last section. This provided a basis to identify the required

knowledge management components and develop system components that would implement the approach. This process also allowed for the reuse of existing code and development work, aligning the process with a wider strategy for simplifying the development of constraint models and development of electronic healthcare records systems using an automated approach. This relied on the development of the Pattern constraint model, which provided a means to implement healthcare record systems according to clinical domain requirements, but also the information governance domain requirements represented by the Secutype model. A series of Pattern editing tools have been developed and supporting infrastructure that generates a web based tool to manage information governance policies.

The design, development and implementation process has been robust and is generalised to the point that the designs could be implemented in a variety of technologies, not just the Enterprise Java Framework. As a proof of concept implementation, Enterprise Java has proven to be very effective in achieving the development of the tool, however the design principles can be extended to other implementations, including potential mobile device and app development. The implementation of *keibi* has been validated as a proof of concept solution to managing information governance requirements, which validates the approach taken to its design and development. The facilities available through the *aruchi* Pattern editing tool provide a means for information governance domain experts to further develop the core Secutype models, whilst the knowledge management framework can update existing systems according to any further developments or implementations. A limitation of this approach is that updates would require a reengineering of the database and adaptation of the old Secutype model into the new schema that would be generated. This is a challenge that remains for all knowledge management solutions, be they in the clinical care or information governance domain. Whilst possible, this approach can be time consuming and complex.

9.5. Evaluating the Knowledge Management Tool

The results gathered during the evaluation sessions represented how a cohort of participants working in the area clinical research interpreted, understood, authored and used information governance policies presented to them by *keibi*. It provided an indication of the kinds of information that they needed and how they put it into practice. Furthermore, the results included a determination of whether participants understood using the tool to author policies, and if they did not, the reasons for their misunderstanding. The results sets also included results relating to the authorship of policies in *keibi*, which are reconciled to the knowledge model framework, and the numbers of responses made when participants use those policy details to answer a series of information handling questions. All of these responses have been tallied with expected responses and a record has been made of where participants behaved consistently with expected outcomes, surpassed expectations, were empowered to use their own expertise or indeed made incorrect decisions, where the causes of error have been defined and specified in each case. Data has also been collected regarding participant views of the system, including the use of a user satisfaction questionnaire to prepare participants for an in depth group discussion about their views of using the system and their opinions regarding the usefulness of the tool.

The purposes of the evaluations were to assess how participants used *keibi* to handle tasks provided to them in exercise sheets. An important metric for considering the usefulness of a knowledge management solution was to discover how participants used the EN ISO 13606 Reference Model to specify policy details, including the number of different types of Class that they used, the quality of details added and reconciling these to see whether there was any effect in allowing participants to answer the information handling questions correctly. This section commences with a discussion and critical appraisal of the results of each experiment, including statistics on the understanding demonstrated by participants, numbers of details added, their accuracy and any effect they had on the participant responses. It continues with a focus on participant experience and views by considering and critically appraising the results of the questionnaire and

group discussions. This leads into a discussion and critique of the model and system architecture.

9.5.1. How Participants Used the System and Models

The analyses presented in section 8.6 provide an overview of the expected uses of *keibi* to author policy excerpts using the Secutype model components. It is clear that all participants were able to use the tool to add and where necessary update policy records without any issues or misunderstanding of the functionality of the tools. It is also clear that participants did not generally follow a particular pattern or expected authoring convention, with a range of numbers of EN ISO 13606 Classes being added across all participants.

This wide range of responses and uses of the knowledge model disproved part of the hypotheses: though understanding was clearly encouraged, it was not consistent across the participants. This inconsistency did not lead to significant misunderstanding when reviewing the authored policy, and the results provided a basis to further develop guidelines around how to express policy using the Secutype model. This also raises the question about whether consistency in understanding is actually desirable, or whether users of a policy should be equipped to handle tasks according to their knowledge and expertise based on a better engagement and education in how to handle information governance requirements.

Participants were given minimal guidance at the beginning of the session because part of the anticipated outcomes of the evaluations was to understand more fully how they might use the knowledge model to complete their tasks given their range of experience across handling clinical information for research purposes, whether they are clinically qualified, purely experienced in research and methodology, responsible for running research IT services, managing governance and policy, or all of the above. In a real usage context, users of *keibi* would be given a training manual and would be supported by a wizard and interactive help screens that were suggested by some participants.

There was little evidence or indication as to how any further training and guidance materials should be prepared: prior to the evaluations there were no

consistent policy framework or detailed specifications on how policy and guidance should be prepared, beyond the core requirements and frameworks proposed by the ISO 27001 and 27002 standards and the various interpretations included those provided by the Department of Health and Information Commissioner's Office in the UK, amongst others. The results of this thesis can help to focus a clearer set of recommended guidelines and approaches through the standardisation of the model now that some metrics and results have been obtained.

A proposed pattern of expectations has been provided as a baseline to compare participant responses. These proposed expectations can not be considered authoritative due to the novel nature of the tool and use of a consistent knowledge model in this case: there are no established or published comparators, no means of independently scrutinising the proposed expectations and no established metrics to represent participant understanding and interpretation. Expectations have been formed on the basis of how the proposed Secutype model allows policy writers and users to focus their attentions on the requirements put before them by legal, procedural, ethical and good practice guidelines and reconciled with the participants' own expected outcomes of the policy excerpt.

The way that participants used the tool is a basis for assessing whether they followed an expected pattern, or whether they added policy items in an entirely different, inconsistent and effective manner. For this reason, it has been important to allow for the acknowledgement that participants exceeded expectations and used their own expertise to guide their responses. This was considered whether the use of their own expertise was assisted by the tool to make correct or incorrect decisions when authoring or using the authored policy. It is arguably hard to exclude other biases in the decision making process for using information security policies and deciding how to express or author them. What the results provide are an indication of how participants tended to proceed, and whether there was a generalisable pattern to their responses. In this case, there was no discernable pattern, and an assessment of the responses in the second and third experiments have provided an indication of whether one is necessary, the effectiveness of the knowledge model and whether this needed to be amended to better support participants, or how they should be guided in using it effectively.

The assessment of how participants used the tool and knowledge model could not proceed with any assumptions about specific knowledge regarding writing of policy, ability to read and interpret policy items, particular expertise with a web application or experience in using policy-based approaches to manage information security. The base assumption for the evaluations is that users and authors of policy can come from any background, are expected to contribute to policy development and adhere to good practice, without any specialist training or education. Were *keibi* used in practice as a commercial system, it is far more likely that the users would be limited to information governance management teams, data managers and / or principal investigators. To that end, the evaluations have provided a more challenging situation than that of adopting environments in working practice. It is a matter of interest that participants generally felt that the tool would be very useful for induction, training and education which is discussed later in the next chapter, however the assumption emphasised the point that the results of how the tool and model were used should be taken as an indication of the use: inferences can be drawn in isolation, but these are limited to what actually got entered and used, and reconciled with participant feedback.

It should be noted that there was arguably a difference in difficulty between the two exercise sheets. Given the range of provided excerpts, the same assumptions were applied to the selection process for the excerpts: quality, accuracy and effectiveness of policy cannot be predicted and there are no baseline considerations to measure these. Additionally, the evaluations have assumed a workflow whereby users would take existing or pre-written policy items and author them using *keibi*, or undocumented, accepted working practice. Whilst it may be possible to take a generic framework (such as ISO 27002) and author policy items based upon that and a knowledge about individual data use context requirements, it has not been possible to evaluate this due to time constraints and providing a representative scenario within a laboratory setting. With further work, the tool should be updated to run within a live test environment, as suggested in the results from experiment three, where these other workflows can be evaluated over time.

A validation of the tool is possible insofar as participants were able to author and use the items correctly and effectively, but also made specific comments on the model itself. These are discussed later in this section, however it should be noted that whilst thematic saturation was achieved with this number of participants within the constraints of the test scenario where limited guidance was given to participants, wider inferences can be drawn on how the tool might be used in a larger, live deployment setting. This work can help to guide how the tool should be deployed and what training can be given, whereupon more data can be gathered about how participants use the tool and whether they feel that they can work effectively with it.

9.5.2. Participant Understanding of Policy Authoring

Appendix 18 and Appendix 19 provide a wealth of results regarding participant understanding, misunderstanding, and correct, incorrect and omitted responses when using the tool to author policies. The results showed some variation in understanding, but some exceeding of expectations when using the tool. The same concerns apply from the last section where expectations were concerned: it is hard to authoritatively specify expected outcomes that can be reliably peer reviewed, and this is one reason why it has been important to include an acknowledgement of participants exceeding expectations.

In general the results for understanding were good – all participants showed more understanding than not, in some cases exceeding expectations. In cases of higher misunderstanding, this was due either to a lack of confidence using the tool, frustration with some of the workflow, or difficulty in translating the detail of the narrative policy excerpts into the appropriate knowledge model Composition. This is a clear indication of the need for training participants to help handle ambiguity in policy specification, though the knowledge management approach does provide a means to focus such training and education on specific tasks rather than make general statements about the need for users to receive appropriate training: the basis and content of that training and education can now be more reliably specified. Additionally, where there is a higher proportion of participant error due to their mistake, this cannot and should not be taken as a reflection of

their abilities – this is purely a means to isolate misunderstanding and supply a reason for that misunderstanding.

Where a participant has shown a misunderstanding as a result of his or her own error, this error can only be attributable to their not being well supported by the policy or it being expressed as a means to support their needs. This itself is supported by half the participants feeling that the tool would provide a personalised resource for individual users, tailored to their specific needs and expected behaviour. It is important to note that misunderstanding due to participant error does not and should not provide a basis to draw conclusions about the competence or abilities of the participants themselves: first, there is no established means of measuring such competency; second, the evaluations are focussed on assessing the tool and knowledge model and not the participants; and third, the tool and model are expected to improve and support users by helping them to see and better understand the requirements and expectations, reducing ambiguity and improving support for users.

It has been clear that users will need specific guidance to author computable policies: whilst two participants were able to author the computable policy provided in question four in both exercise sheets in a way that could inform computable elements within a record system architecture, the rest did not. This suggests that guidance on how to specify these policy items is needed. It should be noted that there were very few examples of computable policy items provided and those that were provided proved too vague for use in this round of evaluations. *keibi* and the knowledge model provide a means of encouraging more specification should original policies be vague.

These results again were at risk of being affected by undetectable biases that are hard to screen. Whilst every effort was made to ensure that participants were comfortable and rested, one at least mentioned that they were quite tired in their questionnaire responses and this might have affected their responses. The categorisation of misunderstanding according to participant error, tool induced error or policy misinterpretation does not take such factors into account, and whilst the participant error category is arguably broad and low, it is not possible to subcategorise any further with reasonable accuracy. Any further exploration of

these factors would require further work, which was beyond the capacity of this thesis.

The results showed that a large number of the responses used the correct Secutype components. Errors were far fewer and generally minor. Omissions, however, were of a higher proportion and participants tended not to use as many of the classes that were expected, particularly Safeguards and Controls. This is due to a deliberately limited guidance and instruction, as well as limited experience of using the tool. Omission scores were based upon the specification of the expected responses, however, which themselves cannot be fully independently validated. The effects of the omissions could however be assessed by the effects on the responses in the second experiment: they did cause some error in the second experiment, but participants tended to use what was available to answer correctly or use their own experience to correctly specify the correct course of action.

9.5.3. Participant Use of Authored Policy Items

The results from the second experiment showed that participants understood the use of *keibi* to guide their behaviour and support their decisions on handling sensitive healthcare information in a research context. The overall majority of answers were correct, with some that surpassed expectations. It was also clear that the participants used their own expertise to override what they read in *keibi* and indicated when they had done this, citing what they saw in the tool and when they thought that their expertise was more appropriate for answering the questions that had been posed.

The determining factors for correct and incorrect answers were based upon what the participants who had offered the original excerpts felt were the expected outcomes of those excerpts. This relied on the participants' ability and understanding of the original policy, as well as the author's assessment of the policy items themselves. Though the outcomes were clear and provided a basis for determining whether participants responded correctly, they do nevertheless depend on interpretation and this, arguably, could be challenged. This is why the responses were scored according to participant experience and use of their own expertise, whether that led to a correct or incorrect response.

The second experiment was also subject to a potentially different level of difficulty between answer sheets. There was no reliable metric for measuring the difficult of individual questions, particularly since some participants had expertise which meant that they could potentially find one question harder than another participant might, and vice versa. This limits the analysis of the results to a representation of what the participants added and how they behaved on using the tool as opposed to comparing any pattern in policy authoring that led to more correct responses, for example. The inferences drawn from these results are also limited to understanding what kinds of information this group of participants found useful and helped them make correct decisions. Further research is needed to find any determining factors for when participants choose to override policy specification in favour of their own expertise and experience. The results nevertheless show that participants did use their own expertise over what was specified in the policy when they used this tool, as well as use what was specified in the tool by their counterparts to successfully answer the majority of the questions that were posed.

9.5.4. Participant Feedback About their Experience Using *keibi*

A widely accepted approach for gathering feedback from participants is the user satisfaction questionnaire. The questionnaire developed by Lewis (James R. Lewis, 1993) is a well tested and validated example, where user opinions can be sought and expanded upon across a scale that has a proven sensitivity for different aspects of the system. The participant sample was too small to run a factor analysis, though the purpose of the satisfaction questionnaire was to gather feedback from the participants and allow them to express their views, providing further insights into the results for the previous experiments and gaining an understanding of any issues that affected use or opinion on the system. Another goal was to prepare participants to engage in a group discussion about their experiences, allowing them to think about and articulate their views, seeing if there were any areas of agreement or different views.

A thematic saturation was reached and the feedback from participants provided valuable insights into the core issues, recommendations and experience

of using the proposed knowledge management approach across a range of role holders within the area of clinical research. The results are limited to participant feedback based upon the model that has been proposed. Whilst participants did not express a desire to specify the model themselves or contribute to the knowledge management framework or design, the evaluations do not provide any insight into how they might respond to doing so. Participants may have felt satisfied that they were able to express what they felt they needed to using the provided model, though some of the feedback suggests that they felt that it was too detailed in some cases, or did not help them to feel productive. The results do not support any indications for changes to the modelling element of the knowledge management approach, other than to show that the approach was considered helpful and held much potential. This suggests further work in the area of exploring user willingness and responses in modelling the Secutype itself prior to deploying a specific application.

9.5.5. Secutype Model

The evaluations have focussed on the effects of the tool on participants and their responses. This approach relies on their experience to provide feedback on the completeness and appropriateness of the model for use. The feedback they provided gave some results regarding the model itself and some validation, which focuses on their opinions. Another more formal approach would be to see whether the tool would be appropriate for assisting with data management plans, section 251 exemption applications, ISO 27001 certification and / or IG Toolkit compliance testing when used in practice, which is a possible area for future investigation. Participants have validated the model in the sense that they felt it was rigorous and complete in their opinion, in some cases the feeling was that it was too detailed. Their suggestion that it and the tool be used for induction, training and education as well as to assist with risk assessment shows that they appreciated the richness of the information that could be stored and how it could be further used to tailor a user specific set of policies to aid understanding and interpretation of expected good practice.

These results focus on a single, modelled representation of security policy and consider participant opinions thereof. This is the only representation that could be provided or developed in line with the requirements that were developed and presented in Chapter 6. Whilst the participants did not suggest any other modelling or approach, there is a possible thread of investigation through further work of how a different representation might work. The focus of participant interest was in part on simplification of the current model, but mainly on useful deployment possibilities and information representation. The proposal from these results is to focus on those areas as opposed to developing different models to represent the requirements that have been specified in Chapter 6.

9.5.6. Use of the Pattern Constraint Model to Implement Secutypes

The use of the Pattern approach has provided the Secutype model in accordance with the EN/ISO 13606 class structure, where information capture requirements, versioning and auditing have been implemented along with a faithful representation of the model itself. This has been used to help develop the web application tool in accordance with the Clinic Manager 3 (C3) framework. Whilst no other knowledge modelling approaches were used or reviewed for the purposes of these evaluations and it is therefore not possible to make claims as to whether this approach is optimal for developing an information governance compliance and decision support tool, the purposes of the evaluations have been to focus on the effects of such an approach on potential users of such a tool.

The Pattern use for implementing the Secutype models has provided a means to establish a proven information structure based upon the ISO standard EN ISO 13606 Reference Model. This has also supplied a means to manage versions of the policy items that are stored within the Secutype model so that edits and deprecation of policies can be managed with full auditing features. The Pattern approach has also meant that deletion of policy items is not directly possible, though errors can be corrected where necessary and a complete record of policy items can be retained even after a project has come to an end or a context of use is otherwise changed or discontinued. The participants identified this as an important feature during the group discussions and one which would be important

to help with the transparency and demonstration of good practice identified in the literature and case studies as being of particular importance for public engagement and involvement.

9.5.7. Clinic Manager 3 Framework

The C3 framework was used to present the user screens and manage the data transfers from the database storage to presentation to the users. The structure has been discussed in Chapter 7, where the Apache Tomcat web server was used to present the screens using the J2EE framework, and the PostgreSQL database was used to persist the data. All software components are used under open source licences, making them an appropriate resource for use in an evaluation environment, and potentially in production. The versions used were stable, and allowed for any required configuration where needed.

The C3 framework provided a pleasing interface for the participants to use, with very simple interface widgets and an understandable convention for screen interaction, not unlike other websites and web applications that the participants were familiar with. The system performed reasonably well, though it did crash unexpectedly during the last evaluation session on 22nd January 2014. The system was being run from a reasonably powered laptop, though this is obviously not a target deployment environment and it is not possible to draw any conclusions from this evaluation about production environment system hardware configurations.

Since a maximum of three participants were using the system at any given time, this will not give any kind of realistic performance metrics or provide any indication of how the current system will scale. This was not the purpose of the evaluations; this work proposes that a suitable production deployment be found for live use analysis. This has been beyond the scope and capability of this research project, however a production environment may have exclusive system requirements and components of the system may need to be altered to work within these environments. The C3 framework has nevertheless been proven successful in getting an application implemented and deployed, providing a range of results that have been gathered in these experiments.

9.6. Strengths and Limitations of the Work

This work has proposed the development of a knowledge management solution to manage the information governance requirements when handling electronic healthcare records for clinical research. It has reviewed requirements sources in the form of legislation, guidelines and international standards, conducted case studies of clinical research projects and developed a series of requirements for a knowledge model and framework to use that model to develop a software tool for handling information governance policies. The tool has been implemented and evaluated in a series of evaluations designed to test the effectiveness of the tool when used by human participants to develop information governance policies and use them to guide their research.

The strengths of the work lie in the first review of information governance sources and literature that includes the legal and ethical bases for protecting confidential information when it is shared for research, including the information security guidelines and standards for developing control structures and advising both human and software actors in the management of research project data holdings. This has guided the specification of the first set of requirements for developing information security policies using a shareable, reusable and common information structure. This has been used to help promote understanding and effective guidance for people when they are handling sensitive healthcare records used for research. It has provided clarity and promoted awareness of what is expected of these people when they handle these records, providing an auditable resource that logs user interactions with policy guidelines and their behaviour with sensitive records. It has developed a method for developing a series of policy documents that can be managed and reviewed by people who are responsible for their management and dissemination.

The development of the Secutype model has provided the first example of an information architecture design for information governance in the area of healthcare record protection and beyond. This includes details that are required for security software configuration as well as the details needed for human readership as established by information security and wider governance policies.

By establishing the model, it has allowed for the structuring and organisation of policy components so that they can be presented meaningfully within given contexts of use. By using a proven information model to represent these concepts, it has been possible to develop a version managed, audited implementation of this structure, and the basis for making further developments in the organisation and management of those components.

This work has tested the approach and implemented tools in a series of evaluations that have been designed to measure its effectiveness when used by live participants as a proof of concept implementation, and to test three hypotheses around the understanding, interpretation and support of human expertise when managing information governance requirements in the context of healthcare information use in clinical research. The approach taken to run the evaluations has used a combination of methods based on core approaches for determining participant understanding, interpretation and behaviour along with an investigation of participant opinions of using the tool for both policy authoring and use to guide them in appropriate behaviour. The author anticipates that this approach can be scaled for larger cohort evaluation in the filed of information governance management where there are limited examples of approaches to evaluate these.

The work has focused on twelve participants in an ethnographic, laboratory based study. The cohort was representative of the skills sets found in the clinical research profession, the evaluations were in a laboratory setting with ideal conditions. This was an ideal approach to provide evidence of the tool as a proof of concept implementation, showing the utility and contribution of the tool and providing a basis for further development before it is released into a live working environment. The results are limited to the laboratory setting and further work will be needed to bring the tool to a point where a pilot live deployment can be possible, and potentially a larger cohort can be selected to evaluate the use of tool over a period of time. It has also not been possible to run the factor analysis on the participant questionnaires, where about one hundred participants would be needed to achieve this and get a clear indication of the factors that are significant for participants when they use the tool. Additionally, the participants selected all

had some association with the School of Life and Medical Sciences at UCL, though worked for different organisations within UCL and the NHS. Future evaluations might include members of other academic institutions.

In conclusion, it is clear that the knowledge management approach encouraged understanding, engagement and sound interpretation of the information governance requirements for experienced members of the clinical research community in a laboratory setting. Overall the participants were positive about the tool and contributed several ideas as to how it could be best deployed in practice. Further evaluation of the approach in live use will provide more evidence to support this work and develop the approach further as working practice and core requirements evolve. The measurement of the effects of the tool will remain difficult whilst there are no accurate or authoritative means to measure success that do not rely upon self reporting of breaches, legal proceedings or monetary fines. The focus of any further evaluations must rely on the proof of understanding and appropriate behaviour as well as clear feedback from the users of the system who are handling the electronic healthcare records for clinical research.

Chapter 10. Further Work and Conclusions

The research work and evaluations have shown that the knowledge management approach encourages understanding and correct interpretation of information governance requirements when managing healthcare information in clinical research. This has involved a set of participants who also gave feedback about their experiences of using the tool. Between the evaluations and the analysis of the results, this work has proven the knowledge management approach and an implementation of it as a proof of concept contribution to managing information governance policy in the clinical research space. The work has also tested three hypotheses, providing results that partially prove them. The research work and evaluations have also provided a source of potential further work. This chapter describes the areas of further work, and concludes the thesis of research.

10.1. Further Work

The further work identified through the research and evaluation results and described in Chapter 9 include two main categories: further Secutype and *keibi* development and wider evaluation potential in live practice. This section describes these areas and proposes additional work for them to come to fruition. One significant activity that should be performed would be to conduct a risk assessment on the use of the Secutype driven solution itself. This would help to develop a strategy to protect it appropriately, given that it is itself an information asset with its own set of vulnerabilities.

10.1.1. Secutype and *keibi* Development

The participants made several recommendations for further development of the tool in answering the questionnaire and during the group discussions, which form the basis of further development and research work. A couple of participants recognised that the Secutype approach and *keibi* provided an opportunity to "tailor" the wealth of information governance policy knowledge stored in the tool to individual users, providing them only what is necessary for them to complete

their tasks and to therefore to reduce the amount that had to be reviewed by them for research projects and departments that they were part of.

Participants felt that there was still too much typing involved with specifying the policy, recommending that pre-specified values, drop down lists and common terms were used to ease this overhead. The recommendations open the possibility of developing an agreed term set for information governance concepts, which in turn could be coded variables. This would provide a basis for developing terminologies and use formal ontology scripting to help specify the relationships between the different concepts that are encapsulated in the Secutype knowledge model. This is potentially an important step as the amount of information and examples of information governance policy increases: participants felt that the ability to index and search for specific policy items in a given context of use would be a powerful and helpful feature.

The author identified that the evolution of the Secutype model to support risk assessments and audits of working practice would be the next step in terms of developing it further. Given the feedback from some participants during the group discussions, the risk assessment development seems to be more pressing. Participants identified that *keibi* would be particularly useful for running risk assessments and analyses, perhaps incorporating the information from the Information Assets and Activities to help the process and inform users about potential risks by applying an appropriate risk scoring mechanism. This would require some updates to the Secutype model, placing the risk assessment and analyses information requirements within the Composition, Cluster and Element information architecture that the EN ISO 13606 Reference Model provides. This would act as a means to store the details of that analysis for further review and update, as well as act as an advisory tool for determining policy items and supporting existing stored policies within *keibi*.

The participants supported the development of *keibi* as an educational and new staff induction tool, where they felt that the rich information sources and process that it applied for authoring and reviewing information governance policies would be very helpful in helping users develop the understanding and skills required for these tasks, as well as how to develop and support good working practice in an on-

going fashion. This suggests a further module for *keibi* to support these goals, and perhaps apply a test on whether the participants had understood what they were doing, adapting the evaluation approach used for the first and second experiments. Such a module would provide a commonly agreed, measurable and insightful basis for assessing user capabilities and develop further assurance that users would be able to behave in the way that was expected when handling electronic healthcare records for research purposes.

There were some issues with the tool itself: the error messages that are part of the C3 Framework were felt to be unintuitive and not helpful and there were a couple of small bugs that remained in the tool. These items can be worked into further development work as a process is developed to incorporate new components and develop a set of error messages that are meaningful in a given context, be it for controlling the sharing of data or running risk assessment and analyses.

10.1.2. Evaluation in Live Practice

Having established the utility of the knowledge management approach in the information governance domain, the results and learning from the work in this thesis provides a basis for further development and evaluation in live practice. These include developing further components for educational purposes, risk management and audit of working practice. Once developed, these could be deployed within a series of live practice scenarios for not only clinical research, but also other secondary use contexts and clinical care. It is clear that the information governance requirements and expectations are, in current practice, based on the same principles for the clinical research and wider secondary use contexts as they are for clinical care purposes. The knowledge management framework that this work has developed is applicable in other use contexts, and evaluation in these other settings is also proposed as further work.

A number of possible approaches could be taken to evaluating the knowledge management framework. These could include use in a live setting for research projects, clinical trials and commissioning services where the various features of the knowledge models and *keibi* tool could be used by various members of

information governance and security management system members. This could be performed across institutions using a larger cohort, perhaps no less than one hundred, so that a user satisfaction factor analysis could be performed. A set of cohorts would be developed to ensure that each of the different proposed modules would be evaluated within different organisations. This would provide an evidence basis to compare the different organisations as well as assess a larger number of potential participants and their responses to using the tool.

These plans are being developed to prepare a research grant application to build a larger team of developers and researchers so that the different uses of the knowledge management for policy development and use, audit and risk management can be considered across a larger cohort. The evaluations can also be adapted to suit the needs of the research questions being posed in each of the main areas of information governance management. This would also allow the collation of more data from user satisfactions questionnaires and factor analysis, across a larger cohort number.

10.2. Conclusions

This thesis has proposed a knowledge management approach to manage the legal, ethical, good practice and information security requirements for sharing electronic healthcare records for research purposes and the use of those records in that research. It has included a literature review and a review of the legislation, guidelines, international standards and good working practice documents for the area of information governance when sharing electronic healthcare records for research purposes. It has considered literature pertaining to the area of information governance and use of EHRs in research as well as care, by considering the use of information technology to develop record servers that permit a consistent view and aim to achieve semantic interoperability between the different systems that are used to provide care and feed the research and other secondary use communities. It has reviewed literature on existing stipulations to maintain participant anonymity, exemptions to the Common Law Duty of Confidentiality that permits reuse of identifying information without participant consent, and the ramifications of Data Protection Legislation. The work has

reviewed available literature and information on funder and international governmental support for EHR reuse in clinical research. It has also considered the international standards for developing systems to communicate EHRs, information security management in general and for healthcare information.

The research work has included observations of working practice, stakeholder interviews and development of research and clinical systems to understand the issues with protecting EHR information when shared for research purposes in a series of clinical research projects, healthcare system and registry development projects. It has used these case studies to further understand the challenges that research is faced with when protecting healthcare information in accordance with legal, ethical and good practice guidelines and standards, the information flows within research and beyond, and its importance to the healthcare providers.

The research work has used the learning outcomes from the literature reviews and case studies to develop the proposed knowledge management solution for protecting healthcare records. This development process has followed the Unified Software Development Process to specify requirements for the knowledge management solution, design the solution based on an analysis of these requirements, develop it into an information model and implement it in accordance with existing communication and interoperability standards within an application that can be used to manage information governance for healthcare information in research. This application and the model have been tested as per the USDP directives, and evaluated in a laboratory setting by human participants to test the effectiveness of the tool, the impact of its use on participant behaviour, and to validate the knowledge model at the centre of the knowledge management solution.

The evaluations and analyses have tested three hypotheses focused on human understanding, interpretation and support to help them meet the information governance requirements and achieve good working practice in line with expectations that have been established in the legal, procedural and practical guidelines. This section concludes the research work, summarising the learning outcomes of the research and contributions made to furthering the understanding of information governance in the area of clinical research.

10.2.1. Challenges for the Information Governance of Healthcare Records

Healthcare information is gathered during care delivery, under an oath of confidentiality that remains the duty of the medical profession to uphold, and is the basis of a relationship of trust between the patent and healthcare provider. This duty and relationship are protected in law within the UK, Europe and beyond. As the use of electronically held clinical and social information to provide care services have increased, the value and benefits of sharing that information more widely have been identified and have informed information management strategies across the UK, Europe and beyond. This has placed more dependency on other legal protections, including data protection and the right to personal privacy, which have been established and evolved to provide assurance about how information gathered during the provision of care is protected in the interests of the patient, their clinical professional and the services that provide that care.

The trend to share information has made the protection of patient confidentiality harder to maintain, placed more expectation on data protection legislation, and caused significant debate around whether the individual rights to personal privacy are being threatened by this sharing. Emphasis on the collection of appropriate consent for the use of information remains, though the practicality and meaningfulness of this consent has become questionable, particularly with the increase in information use for purposes that are secondary to the provision of care, including research. Whilst anonymity is viewed as a means to remain compliant with the legal requirements and protection expectations where consent has not been sought, there are still risks of re-identification despite a number of techniques to de-identify records. The usefulness of information is also reduced by the removal of identifying data.

A series of guidelines and codes of practice have been developed to help people who are responsible for handling and protecting the information as it is used. There remains however a continued pressure to keep abreast of requirements as guidelines and legal frameworks are updated, and this is exacerbated by evidence of poorly understood requirements of the expectations, and a lack of a consistent, clear representation of these requirements across the various contexts of use

within which the information flows occur. Practical guidance is available in the form of internationally acknowledged information security standards and practices, however there remains no clear method of developing an understanding of the legal, ethical and good practice requirements to inform the development of the tools and processes identified by the standards.

There also remains an anxiety over where sensitive, personal information is being sent and how it is being used. This has prompted a series of information governance reviews about how healthcare information is handled for care and beyond. There remain societal concerns about this sharing, which is evident through a series of petitions and commentary about the protection of information and how identifying supposedly de-identified records remain. Recent investigations into data breaches have confirmed anxieties that information has been breached, particularly during processing for secondary activities. This has served as a reminder that there are risks associated with the processing of this information and the care that must be taken with it. This has not stopped the trend for governments to continue to develop the sharing of information to support uses that are secondary to providing care, recognised as being important to maintain services, develop treatment policies and improve care outcomes. There is also a gradual recognition that patients and the public want to know how their information is being used, and are keen to develop an understanding of that information not only for managing their own health, but also for the research that is performed with it.

10.2.2. Sharing Healthcare Records for Research

The sharing of healthcare information for clinical research has been recognised as a compelling reason to break the confidence within which the information was gathered in the first place, either with participant consent or without. Advances in understanding treatment outcomes, patient experience and developing the provision of services to improve those outcomes rely on the availability of information that has been collected under this sworn duty of confidentiality. UK and European governments and research councils have made provision to support this sharing of information, as well as build the facilities and research repositories

available for performing research. The wider sharing for research and other purposes has occurred on the basis of assumptions about the anonymity of the information, requesting permission to set aside the Common Law Duty of Confidentiality or seeking consent where possible. Amendments to acts of law in the UK have provided support for this sharing, though not without controversy.

In combination with the provision of storing and managing healthcare information, development has continued for digitising healthcare records so that they can be meaningfully shared between clinical professionals for care, as well as for research purposes. A series of standardised electronic healthcare record specifications have been developed to create shareable, commonly interpreted and understood models that represent clinical concepts and guide the development of record keeping systems used for providing care. This is in combination with a patient empowerment strategy, where the UK is attempting to grant patients access to their records to help them manage their healthcare more effectively. The standards have been developed using a dual modelling approach, where the structure of the records has been established on the records, and a constraint model specifies instances of that structure to represent clinical concepts in a record. This imposes the common view of the clinical concepts.

By using a series of agreed terms and clinical coding where relationships are formally defined using ontologies, the concepts can be commonly understood between editors and viewers of that information. The combination of these standards and modelling paradigms form the basis for clinical information systems, and together they represent a use of knowledge management to support users of clinical information. The use of knowledge management in the clinical domain has provided an example of how it can bring a consistent view of commonly shared information structures, supporting and developing a consistent and collaborative view of the information so that tasks can be performed effectively.

By adapting this approach, the research work reported in this thesis has included the development of a knowledge management solution for the domain of information governance when using electronic healthcare records for research. The author has theorised that the facilities for sharing healthcare records can be

used to share necessary information about how to protect this information as it is used for research purposes. This includes sharing a common understanding and interpretation of those protection measures between the people who are responsible for developing policy based controls and those who are responsible for working safely and responsibly with healthcare records.

10.2.3. Why Information Governance Needs Knowledge Management

The complexity of the legal, procedural and good practice expectations for protecting healthcare information during research uses has caused a lack of clarity and limited understanding for people responsible for conducting research, managing the flow of information from clinical care into research repositories and the repositories that manage that information. This is exacerbated by a slow process of developing narrative policy documents, codes of practices and sharing agreements that have to be read, understood and put into practice either directly by people, or indirectly by refining the requirements to a software and computer processed set of heuristics that manage access controls, privileges and data de-identification. There are limited means to log the development process and use of these controls, and maintaining transparency and demonstrating good practice is hard. This is occurring during a period of increased public anxiety over the use of personally identifying, confidential healthcare records, where a legal position regarding data protection is being developed that defines more stringent requirements for obtaining consent and there is a wider expectation of transparency and accountability for the use of these healthcare records.

By applying a knowledge management solution to this problem, the author theorised that it would clarify the information governance requirements for users, simplify the process of developing policy based controls, encourage a consistent understanding and interpretation of what was required and support user expertise when handling healthcare record information whilst conducting research. It would also provide a basis to show the evolution and development of these information governance policies and how they are used in practice, to the point that a research project would be able to show how it had developed control mechanisms and that the people involved had seen and read the policy items.

These features were expected to provide the basis for an audit of procedure and practice. The author developed a knowledge management solution that would test this theory.

10.2.4. Development and Evaluation of the Knowledge Management Solution

By applying the widely used Unified Software Development Process, the author has gathered requirements, designed, developed, implemented and tested a candidate knowledge management framework and evaluated it as a proof of concept solution for managing information governance requirements. This entailed the design and development of a new information governance knowledge model, facilities to author instances of this model, and a framework to generate a policy management tool based on this information model. Once implemented, the model and tool were tested in a series of pilot evaluations, where amendments were made to the model and tool for further use in experiments that were designed to evaluate the approach.

The thesis proposed the tool as a proof of concept implementation and to test three hypotheses. The hypotheses stated that the knowledge management approach would encourage a consistent understanding of expected behaviour across a range of role holders when authoring and reviewing information governance policy; limit variation in interpreting information governance requirements when authoring and reviewing policies; and support user expertise when interpreting required behaviour and refined these requirements to computable heuristics. The evaluations involved live participants and showed that the approach encouraged understanding though this was not always consistent across the exercises, that the interpretations were not always consistent but that user expertise was supported and that overall they were able to make correct, meaningful decisions about how to author and use policies using the tool.

The approach was also evaluated by seeking participant feedback on using the tool and asking them to discuss their experience, assessing whether they found it useful and if they would use it in practice. The responses provided a wealth of feedback relating to the use of the framework and tool, which showed that participants felt that it offered a better solution than what existed to support them

currently, they understood the tool and the knowledge model and found it helpful. They also provided proposals for further areas of investigation, including the use of the tool for educational purposes and a means to test users on their knowledge and abilities with managing healthcare records.

There was unanimous support for developing "tailored" policies that showed users what they needed to see in the wealth of information that would be stored and organised by the policy tool. There was also a strong desire for wizards to help with the policy authoring and interactive help screens to assist this process. The evidence clearly showed that, in order to encourage a more consistent understanding and interpretation of what was required, users would need to be guided clearly and effectively. This was especially clear if the stored data was to be refined to computable heuristics. It was also clear that it was important to develop a common agreement on the intentions of the policy items so that success criteria can be meaningfully established and that the wealth of user expertise and tacit domain knowledge can be captured to reinforce future evaluation frameworks. By adapting this approach, the author has proposed further work to consider the areas of risk assessment as supported by participants feedback, as well as a possibility of developing professional skills using the tool and evaluation framework used in this thesis as marking criteria.

In conclusion, the knowledge management approach has provided a basis for helping to manage information governance requirements when sharing healthcare information in clinical research, and managing its protection as it is used. It has provided a basis for a powerful, usable tool that eases the complexity of these requirements for users, provides a basis for computable refinement and opens up areas of other research interest, particularly in terms of developing the framework for use in auditable risk management, working practice audit and developing training and educational materials.

10.2.5. Closing Remarks

This thesis has provided an insight into the bases for using information collected when people present for healthcare services to support clinical research and other secondary purposes. The work has highlighted the support for reusing that

information for this variety of purposes, one of the most compelling being clinical research, which is governed by stringent legal and ethical requirements as well as the need to maintain research excellence and high standards for meaningful results to be obtained that will help to improve healthcare service provision and outcomes for the people who present for care. To achieve this effectively and safely, the research community must work to high standards that protect the individual's right to privacy and uphold the solemn duty of confidentiality, the basis of a relationship of trust between patient and practitioner.

The focus of the efforts must therefore be on people: those who present for care, those who use the records of that care for research, and those who are involved with both. People must be engaged so that they understand what is expected of them in performing research and protecting the individuals about whom the records have been kept, whilst the wider public must be engaged so that they understand why research is important, and why they are sometimes being asked to allow others to see the information outside of that trusting relationship with their clinical professional. Without understanding the risks involved, which include the likelihood of re-identification, the nature of the research and whether it is likely to benefit them or society more widely, they cannot hope to be able to make meaningful decisions about what it means to them and whether they wish to participate. It is clearly not sufficient to assume that record holders have a mandate to share without engaging with the individual, regardless of the scale of their study or perceived benefits.

Any meaningful discourse about these matters relies on transparency and effective management of the risks and expectations of society, which can be used to provide assurance that records are being handled with due respect, care and diligence. A knowledge management approach has been shown to help make these requirements clearer and encourage understanding and consideration about information governance. It has also been shown to be a meaningful basis for education in the area so that people comply with policy items and understand why they are being asked to do so. The features provide a means for the public to understand how their information is protected and a basis for a meaningful engagement in the area of protection.

There is still work to be done so that a meaningful discourse and rigorous assurance can be provided to the public. The knowledge management approach offers a foundation for this. This is because it is a means to engage people, helping them to understand how to be successful in working to reduce risks and becoming part of their professional experience and working practice. Security is best applied when it does not hinder reasonable working practice: to that end, knowledge management is a way of supporting the evolution of working practice to include effective information governance as an enabler, not a hindrance. If the international strategies for enabling better healthcare through effective use of information technology are to succeed, they cannot be hindered by lack of engagement, misunderstanding and the resultant exclusion of the people who are the intended recipients of these improvements.

Chapter 11. Bibliography

AAMOT, H., KOHL, C. D., RICHTER, D. & KNAUP-GREGORI, P. 2013. Pseudonymization of patient identifiers for translational research. *BMC Med Inform Decis Mak,* 13**,** 75.

ACADEMY OF MEDICAL SCIENCES 2006. Personal data for public good: using health information in medical research. Academy of Medical Sciences.

AGRAWAL, R. & JOHNSON, C. 2007. Securing electronic health records without impeding the flow of information. *Int J Med Inform,* 76**,** 471-9.

AINSWORTH, J. & HARPER, R. 2007. The PsyGrid Experience: Using Web Services in the Study of Schizophrenia. *International Journal of Healthcare Information Systems and Informatics (IJHISI),* 2**,** 1-20.

AL-FEDAGHI, S. S. 2007. Beyond purpose-based privacy access control. *Proceedings of the eighteenth conference on Australasian database - Volume 63.* Ballarat, Victoria, Australia: Australian Computer Society, Inc.

AL-SHAHI, R., VOUSDEN, C. & WARLOW, C. 2005. Bias from requiring explicit consent from all participants in observational research: prospective, population based study. *British Medical Journal,* 331**,** 942 - 946.

AL-SHAHI SALMAN, R., BELLER, E., KAGAN, J., HEMMINKI, E., PHILLIPS, R. S., SAVULESCU, J., MACLEOD, M., WISELY, J. & CHALMERS, I. 2014. Increasing value and reducing waste in biomedical research regulation and management. *Lancet,* 383**,** 176-85.

ALHAQBANI, B. & FIDGE, C. 2008. Access Control Requirements for Processing Electronic Health Records. *In:* HOFSTEDE, A., BENATALLAH, B. & PAIK, H.-Y. (eds.) *Business Process Management Workshops.* Springer Berlin Heidelberg.

ANDERSON, J. G. 2000. Security of the distributed electronic patient record: a case-based approach to identifying policy issues. *Int J Med Inform,* 60**,** 111-8.

ANDERSON, J. G. 2007. Social, ethical and legal barriers to e-health. *Int J Med Inform,* 76**,** 480-3.

ANDERSON, R., BROWN, I., DOWTY, T., INGLESANT, P., HEATH, W. & SASSE, A. 2009. Database State - A Report Commissioned by the Joseph Rowntree Reform Trust.

ANGELL, E., SUTTON, A. J., WINDRIDGE, K. & DIXON-WOODS, M. 2006. Consistency in decision making by research ethics committees: a controlled comparison. *J Med Ethics,* 32**,** 662-4.

ANGUS, V. C., ENTWISTLE, V. A., EMSLIE, M. J., WALKER, K. A. & ANDREW, J. E. 2003. The requirement for prior consent to participate on survey response rates: a population-based survey in Grampian. *BMC Health Serv Res,* 3**,** 21.

ATHEY, B. D., BRAXENTHALER, M., HAAS, M. & GUO, Y. 2013. tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. *AMIA Jt Summits Transl Sci Proc,* 2013**,** 6-8.

AUSTIN, T. 2004. *Development of an Electronic Healthcare Record Architecture utilising Distributed Objects and Directory Services with the Example of Cardiovascular Disease Systems.* PhD PhD, UCL.

AUSTIN, T., KALRA, D., LEA, N. C., PATTERSON, D. L. & INGRAM, D. 2009. Analysis of Clinical Record Data for Anticoagulation Management within an EHR System. *Open Med Inform J,* 3**,** 54-64.

AUSTIN, T., SUN, S., HASSAN, T. & KALRA, D. 2013. Evaluation of ISO EN 13606 as a result of its implementation in XML. *Health Informatics J,* 19**,** 264-80.

BAND, K. 2014. *NHS Care.data: Our Meeting with NHS England Yesterday* [Online]. 38 Degrees. Available: http://blog.38degrees.org.uk/2014/02/19/nhs-care-data-our-meeting-with-nhs-england-yesterday/ [Accessed 27th October 2014].

BARR, M., SOUAN, L., MACGABHANN, P., MULLER, J., AL ASHHAB, M., JASSER, M., HAMZA, K., AL HASSOON, S., KUHN, U., INFANTE, D., LAWLOR, D., GATELY, K., AMIREH, E., O'BYRNE, K. & SUGHAYER, M. A. 2014. The establishment of an ISO compliant cancer biobank for Jordan and its neighboring countries through knowledge transfer and training. *Biopreserv Biobank,* 12**,** 3-12.

BARRETT, G., CASSELL, J. A., PEACOCK, J. L. & COLEMAN, M. P. 2006. National survey of British public's views on use of identifiable medical data by the National Cancer Registry. *Bmj,* 332**,** 1068-72.

BAU, C. T., CHEN, R. C. & HUANG, C. Y. 2014. Construction of a clinical decision support system for undergoing surgery based on domain ontology and rules reasoning. *Telemed J E Health,* 20**,** 460-72.

BEALE, T. 2002. Archetypes: Constraint-Based Domain Models for Future-Proof Information Systems. *In:* K. BACLAWSKI, H. K. (ed.) *Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer.* Seattle, Washington, USA.

BECKER, M. 2005. *Cassandra: Flexible Trust Management and its Application to Electronic Health Records* [Online]. Microsoft Reseaarch. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=76079 [Accessed 27th October 2014].

BECKER, M. 2007. Information governance in NHS's NPfIT: A case for policy specification. *International Journal of Medical Informatics,* 76**,** 432 - 437.

BECKER, M., MALKIS, A. & BUSSARD, L. 2010. *S4P: A Generic Language for Specifying Privacy Preferences and Policies* [Online]. Microsoft Research. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=122108 [Accessed 27th October 2014].

BERGES, I., BERMUDEZ, J. & ILLARRAMENDI, A. 2014. Binding SNOMED CT Terms to Archetype Elements. Establishing a Baseline of Results. *Methods Inf Med,* 53.

BERNAL, J. G., LOPEZ, D. M. & BLOBEL, B. 2012. Architectural approach for semantic EHR systems development based on Detailed Clinical Models. *Stud Health Technol Inform,* 177**,** 164-9.

BLOBEL, B. 2002. Comparing concepts for electronic health record architectures. *Stud Health Technol Inform,* 90**,** 209-14.

BLOBEL, B. 2004. Authorisation and access control for electronic health record systems. *Int J Med Inform,* 73**,** 251-7.

BLOBEL, B. 2006. Advanced and secure architectural EHR approaches. *Int J Med Inform,* 75**,** 185-90.

BLOBEL, B. 2007. Comparing approaches for advanced e-health security infrastructures. *International Journal of Medical Informatics / Virtual Biomedical Universities and E-Learning and Secure eHealth: Managing Risk to Patient Data - E-Learning and Secure eHealth Double* 76**,** 454-459.

BLOOM, B. H. 1970. Space/time trade-offs in hash coding with allowable errors. *Commun. ACM,* 13**,** 422-426.

BOBROW, M. 2013. Balancing privacy with public benefit. *Nature,* 500**,** 123.

BOINTNER, K. & DUFTSCHMID, G. 2009. HL7 template model and EN/ISO 13606 archetype object model - a comparison. *Stud Health Technol Inform,* 150**,** 249.

BOYNTON, P. M. 2005. *The Research Companion: A Practical Guide for the Social and Health Sciences*, Psychology Press.

BRITISH COMPUTER SOCIETY. 2014. *British Computer Society Latest Security Articles* [Online]. BCS. Available: http://www.bcs.org/category/11307 [Accessed 27th October 2014].

BRITISH STANDARDS INSTITUTE 2005a. ISO IEC 27002: 2005 Information technology -- Security techniques -- Code of practice for information security management. British Standards Institute.

BRITISH STANDARDS INSTITUTE 2005b. ISO/IEC 27001: 2005 Information technology -- Security techniques -- Information security management systems -- Requirements. British Standards Institute.

BRITISH STANDARDS INSTITUTE 2007a. BS EN 13606-1: 2007 Health informatics — Electronic health record communication — Part 1: Reference model.

BRITISH STANDARDS INSTITUTE 2007b. BS EN 13606-2: 2007 Health informatics — Electronic health record communication — Part 2: Archetypes interchange specification.

BRITISH STANDARDS INSTITUTE 2009. BS EN 13606-4: 2009 Health informatics -- Electronic health record communication -- Part 4: Security.

BRITISH STANDARDS INSTITUTE 2013a. BS ISO/IEC 27002: 2013 Information technology - Security techniques - Code of practice for information security controls. British Standards Institute.

BRITISH STANDARDS INSTITUTE 2013b. ISO/IEC 27001: 2013 Information technology -- Security techniques -- Information security management systems -- Requirements. BSI.

CAINE, K. & HANANIA, R. 2013. Patients want granular privacy control over health information in electronic medical records. *J Am Med Inform Assoc,* 20**,** 7-15.

CALLWAY, E. 2013. UK Push to Open Up Patients' Data. *Nature,* 502.

CARRION SENOR, I., FERNANDEZ ALEMAN, J. L. & TOVAL, A. 2012. [Access control management in electronic health records: a systematic literature review]. *Gac Sanit,* 26**,** 463-8.

CAULFIELD, T. & ZARZECZNY, A. 2014. Defining 'medical necessity' in an age of personalised medicine: A view from Canada. *Bioessays,* 36**,** 813-7.

CHADWICK, D. W. & OTENKO, A. 2003. The PERMIS X.509 role based privilege management infrastructure. *Future Generation Computer Systems,* 19**,** 277-289.

CHALMERS, J. & MUIR, R. 2003. Patient privacy and confidentiality. *British Medical Journal (BMJ),* 326**,** 725-726.

CHEN, H., CHEN, X., GU, P., WU, Z. & YU, T. 2014. OWL Reasoning Framework over Big Biological Knowledge Network. *Biomed Res Int,* 2014**,** 272915.

CHIME EHR GROUP. 2014. *The UCL Website: Welcome to EHR / CHIIME Ventures* [Online]. Available: http://www.ehr.chime.ucl.ac.uk [Accessed 27th October 2014].

CLINICAL PRACTICE RESEARCH DATALINK. 2014. *Clinical Practice Research Datalink Website: Welcome to the Clinical Pratcice Research Datalink* [Online]. Available: http://www.cprd.com/home/ [Accessed 27th October 2014].

COLLABORATIVE HIV PAEDIATRIC STUDY. 2014. *Collaborative HIV Paediatric Study Website* [Online]. Available: http://www.chipscohort.ac.uk/default.asp [Accessed 27th October 2014].

CONNECTING FOR HEALTH. 2011. *Connecting for Health: Information Governance* [Online]. Available: http://www.connectingforhealth.nhs.uk/systemsandservices/infogov [Accessed 06/04/2011 2011].

COOKSEY, S. D. 2006. A review of Health Service Funding. United Kingdom.

CORRIGAN, D., HEDERMAN, L., KHAN, H., TAWEEL, A., KOSTOPOULOU, O., DELANEY, B., CORRIGAN, D., HEDERMAN, L., KHAN, H., TAWEEL, A., KOSTOPOULOU, O. & DELANEY, B. 2013. An Ontology-Driven Approach to Clinical Evidence Modelling Implementing Clinical Prediction Rules. *In:*

KASTANIA, A. & KASTANIA, A. (eds.) *E-Health Technologies and Improving Patient Safety.* Hershey, PA.

CORRIGAN, O. 2003. Empty ethics: the problem with informed consent. *Sociology of Health & Illness,* 25**,** 768-792.

DANIEL, C. & CHOQUET, R. 2014. Information technology for clinical, translational and comparative effectiveness research. Findings from the section clinical research informatics. *Yearb Med Inform,* 9**,** 224-7.

DE LUSIGNAN, S. 2014. Effective pseudonymisation and explicit statements of public interest to ensure the benefits of sharing health data for research, quality improvement and health service management outweigh the risks. *Inform Prim Care,* 21**,** 61-3.

DE LUSIGNAN, S., CHAN, T., THEADOM, A. & DHOUL, N. 2007. The roles of policy and professionalism in the protection of processed clinical data: A literature review. *International Journal of Medical Informatics,* 76**,** 261-268.

DE LUSIGNAN, S., SULLIVAN, F. & KRAUSE, P. 2010. Vault, cloud and agent: choosing strategies for quality improvement and research based on routinely collected health data. *Inform Prim Care,* 18**,** 1-4.

DEKKER, M. A. C. & ETALLE, S. 2007. Audit-Based Access Control for Electronic Health Records. *Electron. Notes Theor. Comput. Sci.,* 168**,** 221-236.

DELANEY, B. 2009. Evidence-based diagnosis in general practice: needs both robust evidence and sophisticated electronic health record systems. *Family Practice,* 26**,** 239-240.

DELANEY, B. C., PETERSON, K. A., SPEEDIE, S., TAWEEL, A., ARVANITIS, T. N., HOBBS, F. D. R., DELANEY, B. C., PETERSON, K. A., SPEEDIE, S., TAWEEL, A., ARVANITIS, T. N. & HOBBS, F. D. R. 2012. Envisioning a Learning Health Care System: The Electronic Primary Care Research Network, A Case Study. *ANNALS OF FAMILY MEDICINE,* 10**,** 54-59.

DEPARTMENT FOR BUSINESS, I. A. S. 2014. *Department for Business Innovation and Skills Website* [Online]. Her Majesty's Stationary Office. Available: http://www.bis.gov.uk/ [Accessed 27th October 2014].

DEPARTMENT OF HEALTH 2001. Building the Information Core – Implementing the NHS Plan. Department of Health.

DEPARTMENT OF HEALTH. 2003. *Confidentiality: NHS Code of Practice* [Online]. Department of Health. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/200146/Confidentiality_-_NHS_Code_of_Practice.pdf [Accessed 27th October 2014].

DEPARTMENT OF HEALTH. 2006. *Records Management: Code of Practice Parts 1 and 2* [Online]. Department of Health. Available: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4131747 [Accessed 27th October 2014].

DEPARTMENT OF HEALTH. 2007. *Information Security Management: Code of Practice* [Online]. Department of Health. Available: http://systems.hscic.gov.uk/infogov/codes/securitycode.pdf [Accessed 04/10/2014 2014].

DEPARTMENT OF HEALTH. 2010a. *Caldicott Guardian Manual* [Online]. Department of Health. Available:

http://systems.hscic.gov.uk/infogov/links/2010cgmanual.pdf [Accessed 27th October 2014].

DEPARTMENT OF HEALTH. 2010b. *Confidentiality: NHS Code of Practice Supplementary Guidance: Public Interest Disclosures* [Online]. Department of Health. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/200147/Confidentiality_-_NHS_Code_of_Practice_Supplementary_Guidance_on_Public_Interest_Disclosures.pdf [Accessed 27th October 2014].

DEPARTMENT OF HEALTH. 2011. *Connecting for Health* [Online]. NHS. Available: http://www.connectingforhealth.nhs.uk/ [Accessed 27th October 2014].

DEPARTMENT OF HEALTH. 2013a. *Care.data Website* [Online]. UK. Available: http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Pages/care-data.aspx [Accessed 27th October 2014].

DEPARTMENT OF HEALTH 2013b. Information: To share or not to share? The Information Governance Review. Department of Health.

DEPARTMENT OF HEALTH. 2014a. *About the Information Governance Toolkit* [Online]. Department of Health. Available: https://www.igt.hscic.gov.uk/resources/About the IG Toolkit.pdf [Accessed 27th October 2014].

DEPARTMENT OF HEALTH. 2014b. *Department of Health Website* [Online]. Department of Health. Available: https://www.gov.uk/government/organisations/department-of-health [Accessed 27th October 2014].

DUFTSCHMID, G., RINNER, C., KOHLER, M., HUEBNER-BLODER, G., SABOOR, S. & AMMENWERTH, E. 2013. The EHR-ARCHE project: satisfying clinical information needs in a Shared Electronic Health Record system based on IHE XDS and Archetypes. *Int J Med Inform,* 82**,** 1195-207.

DUFTSCHMID, G., WRBA, T. & RINNER, C. 2010. Extraction of standardized archetyped data from Electronic Health Record systems based on the Entity-Attribute-Value Model. *Int J Med Inform,* 79**,** 585-97.

DURHAM, E. A., KANTARCIOGLU, M., XUE, Y., TOTH, C., KUZU, M. & MALIN, B. 2013. Composite Bloom Filters for Secure Record Linkage. *IEEE Transactions on Knowledge and Data Engineering,* 99**,** 1-1.

ECONOMIC AND SOCIAL RESEARCH COUNCIL. 2014a. *Administrative Data Taskforce (ADT) Website* [Online]. Available: http://www.esrc.ac.uk/collaboration/collaborative-research/adt/index.aspx [Accessed 27th October 2014].

ECONOMIC AND SOCIAL RESEARCH COUNCIL. 2014b. *Big Data Network Website* [Online]. Available: http://www.esrc.ac.uk/research/major-investments/Big-Data/ [Accessed 27th October 2014].

EHR4CR CONSORTIUM. 2014. *EHR4 General Information Website* [Online]. Available: http://www.ehr4cr.eu/ [Accessed 27th October 2014].

EL EMAM, K. & DANKAR, F. K. 2008. Protecting privacy using k-anonymity. *J Am Med Inform Assoc,* 15**,** 627-37.

ELLIOT, M., PURDAM, K. & SMITH, D. 2006. Statistical disclosure control architectures for patient records in biomedical information systems. *Journal of Biomedical Informatics,* 41**,** 58-64.

EN 13606 ASSOCIATION. 2014. *The CEN/ISO EN13606 standard* [Online]. Available: http://www.en13606.org/the-ceniso-en13606-standard [Accessed 27th October 2014].

EUROPEAN COMMISSION. 2012a. *European Commission Website: Commission proposes a comprehensive reform of the data protection rules* [Online]. Available: http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm [Accessed 27th October 2014].

EUROPEAN COMMISSION. 2012b. *European Commission Website: Research and Innovation FP7* [Online]. European Commission. Available: http://ec.europa.eu/research/fp7/index_en.cfm [Accessed 27th October 2014].

EUROPEAN COMMITTEE FOR STANDARDISATION 2004. CEN TS 14796 - Health Informatics - Data Types.

EUROPEAN MEDICAL INFORMATION FRAMEWORK (EMIF) CONSORTIUM. 2014. *EMIF Website: About EMIF* [Online]. EMIF Consortium. Available: http://www.emif.eu/emif/about-emif [Accessed 27th October 2014].

EUROPEAN PARLIAMENT AND COUNCIL 1995. Council Directive (EC) on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *In:* COUNCIL, E. P. A. (ed.) *95/46/EC.* European Parliament and Commission.

EXETER, D. J., RODGERS, S. & SABEL, C. E. 2014. "Whose data is it anyway?" The implications of putting small area-level health and social data online. *Health Policy,* 114**,** 88-96.

FARFAN SEDANO, F. J., TERRON CUADRADO, M., CASTELLANOS CLEMENTE, Y., SERRANO BALAZOTE, P., MONER CANO, D. & ROBLES VIEJO, M. 2011. Patient Summary and medicines reconciliation: application of the ISO/CEN EN 13606 standard in clinical practice. *Stud Health Technol Inform,* 166**,** 189-96.

FARR INSTITUTE FOR HEALTH INFORMATICS RESEARCH. 2014. *Farr Institute Website: Innovative Governance* [Online]. Available: http://www.farrinstitute.org/85_Innovative-Governance.html [Accessed 27th October 2014].

FARR INSTITUTE OF HEALTH INFORMATICS RESEARCH. 2014. *Farr Institute Website: About the Farr Institute* [Online]. Available: http://www.farrinstitute.org/1_about.html [Accessed 27th October 2014].

FERNANDES, A. C., CLOETE, D., BROADBENT, M. T., HAYES, R. D., CHANG, C. K., JACKSON, R. G., ROBERTS, A., TSANG, J., SONCUL, M., LIEBSCHER, J., STEWART, R. & CALLARD, F. 2013. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. *BMC Med Inform Decis Mak,* 13**,** 71.

FERNANDEZ-ALEMAN, J. L., SENOR, I. C., LOZOYA, P. A. & TOVAL, A. 2013. Security and privacy in electronic health records: a systematic literature review. *J Biomed Inform,* 46**,** 541-62.

FERREIRA, A., CRUZ-CORREIA, R., ANTUNES, L. & CHADWICK, D. 2007. Access control: how can it improve patients' healthcare? *Studies in Health Technology and Informatics,* 127**,** 65-76.

FOSTER, C., JUDD, A., TOOKEY, P., TUDOR-WILLIAMS, G., DUNN, D., SHINGADIA, D., BUTLER, K., SHARLAND, M., GIBB, D. & LYALL, H. 2009. Young people in the

United Kingdom and Ireland with perinatally acquired HIV: the pediatric legacy for adult services. *AIDS Patient Care STDS,* 23**,** 159-66.

GAJANAYAKE, R., IANNELLA, R. & SAHAMA, T. R. 2012. Privacy oriented access control for electronic health records. *Data Usage Management on the Web Workshop at the Worldwide Web Conference.* Lyon Convention Centre, Lyon, France: ACM.

GANSLANDT, T., MATE, S., HELBING, K., SAX, U. & PROKOSCH, H. U. 2011. Unlocking Data for Clinical Research - The German i2b2 Experience. *Appl Clin Inform,* 2**,** 116-27.

GARDE, S., CHEN, R., LESLIE, H., BEALE, T., MCNICOLL, I. & HEARD, S. 2009. Archetype-based knowledge management for semantic interoperability of electronic health records. *Stud Health Technol Inform,* 150**,** 1007-11.

GARDE, S., HOVENGA, E., BUCK, J. & KNAUP, P. 2007a. Expressing Clinical Data Sets with openEHR Archetypes: A Solid Basis for Ubiquitous Computing. *International Journal of Medical Informatics* 76**,** 334-341.

GARDE, S., KNAUP, P., HOVENGA, E. & HEARD, S. 2007b. Towards semantic interoperability for electronic health records. *Methods Inf Med,* 46**,** 332-43.

GENOMICS ENGLAND. 2013. *Genomics England Website: Genomics England and the 100,000 Genomes Project* [Online]. Department of Health. Available: http://www.genomicsengland.co.uk/wp-content/uploads/2014/07/Narrative-Genomics-England-the-100000-Genomes-Project-FINAL-28-7-14.pdf [Accessed 27th October 2014].

GOOSSEN, W., GOOSSEN-BAREMANS, A. & VAN DER ZEL, M. 2010. Detailed clinical models: a review. *Healthc Inform Res,* 16**,** 201-14.

GOOSSEN, W. T. 2008. Using detailed clinical models to bridge the gap between clinicians and HIT. *Stud Health Technol Inform,* 141**,** 3-10.

GOTTESMAN, O., KUIVANIEMI, H., TROMP, G., FAUCETT, W. A., LI, R., MANOLIO, T. A., SANDERSON, S. C., KANNRY, J., ZINBERG, R., BASFORD, M. A., BRILLIANT, M., CAREY, D. J., CHISHOLM, R. L., CHUTE, C. G., CONNOLLY, J. J., CROSSLIN, D., DENNY, J. C., GALLEGO, C. J., HAINES, J. L., HAKONARSON, H., HARLEY, J., JARVIK, G. P., KOHANE, I., KULLO, I. J., LARSON, E. B., MCCARTY, C., RITCHIE, M. D., RODEN, D. M., SMITH, M. E., BOTTINGER, E. P. & WILLIAMS, M. S. 2013. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med,* 15**,** 761-71.

GRIFFON, N., CHARLET, J. & DARMONI, S. 2013. Knowledge representation and management: towards an integration of a semantic web in daily health practice. *Yearb Med Inform,* 8**,** 155-8.

HAGA, S. B. & O'DANIEL, J. 2011. Public perspectives regarding data-sharing practices in genomics research. *Public Health Genomics,* 14**,** 319-24.

HANNAN, A. 2010. Providing patients online access to their primary care computerised medical records: a case study of sharing and caring. *Inform Prim Care,* 18**,** 41-9.

HANNAN, A. & WEBBER, F. 2007. Towards a partnership of trust. *Stud Health Technol Inform,* 127**,** 108-16.

HAYCOX, A., PIRMOHAMED, M., MCLEOD, C., HOUTEN, R. & RICHARDS, S. 2014. Through a Glass Darkly: Economics and Personalised Medicine. *Pharmacoeconomics*.

HEALTH AND SOCIAL CARE INFORMATION CENTRE. 2014a. *The Health and Social Care Information Centre Website* [Online]. Available: http://www.hscic.gov.uk/ [Accessed 27th October 2014].

HEALTH AND SOCIAL CARE INFORMATION CENTRE. 2014b. *HSCIC Website: Hospital Episode Statistics* [Online]. Available: http://www.hscic.gov.uk/hes [Accessed 27th October 2014].

HEALTH AND SOCIAL CARE INFORMATION CENTRE. 2014c. *HSCIC Website: NHS Number* [Online]. HSCIC. Available: http://systems.hscic.gov.uk/nhsnumber [Accessed 27th October 2014].

HEALTH AND SOCIAL CARE INFORMATION CENTRE. 2014d. *Secondary Uses Service Website* [Online]. Available: http://www.hscic.gov.uk/sus [Accessed 27th October 2014].

HEALTH AND SOCIAL CARE INFORMATION CENTRE (HSCIC). 2014a. *HSCIC Website: Information Governance (IG)* [Online]. Available: http://systems.hscic.gov.uk/infogov [Accessed 27th October 2014].

HEALTH AND SOCIAL CARE INFORMATION CENTRE (HSCIC). 2014b. *HSCIC Website: What Are My Choices?* [Online]. NHS. Available: http://systems.hscic.gov.uk/scr/patients/what/index_html [Accessed 27th October 2014].

HEALTH AND SOCIAL CARE INFORMATION CENTRE HSCIC. 2013. *HSCIC Website: Accredited Safe Havens* [Online]. Available: http://www.hscic.gov.uk/media/12203/Accredited-Safe-Haven-Accreditation-Process-Stage-1---June-2013/pdf/safe-haven-accred-proc-stage-1.pdf [Accessed 27th October 2014].

HEALTH LEVEL 7 (HL7). 2014a. *The Health Level 7 (HL7) Website: Introduction to HL7 Standards* [Online]. Available: http://www.hl7.org/implement/standards/index.cfm?ref=nav [Accessed 27th October 2014].

HEALTH LEVEL 7 (HL7). 2014b. *HL 7 Website: HL7 Templates* [Online]. Available: http://www.hl7.org/Special/committees/template/index.cfm [Accessed 27th October 2014].

HEALTH RESEARCH AUTHORITY. 2014a. *HRA Website: About the Health Research Authority* [Online]. Health Research Athority. Available: http://www.hra.nhs.uk/about-the-hra/ [Accessed 27th October 2014].

HEALTH RESEARCH AUTHORITY. 2014b. *HRA Website: Data Legislation and Information Governance* [Online]. Available: http://www.hra.nhs.uk/resources/data-legislation-and-information-governance/ [Accessed 27th October 2014].

HEALTHIMAGING 2008. EU court orders Finland to pay fine for employee medical data breach. *HealthImaging.* TriMED Media Group.

HEMMINKI, E. 2005. Research ethics committees: agents of research policy? *Health Res Policy Syst,* 3**,** 6.

HER MAJESTY'S SATIONARY OFFICE 2012. The power of information: Putting all of us in control of the health and care information we need. London: Department of Health.

HER MAJESTY'S SATIONARY OFFICE 2014. Care Act 2014. *In:* OFFICE, H. M. S. S. (ed.). London: Her Majesty's Stationary Office.

HER MAJESTY'S SATIONARY OFFICE (HMSO) 1998. Human Rights Act 1998. HMSO London: Her Majesty's Stationary Office.

HER MAJESTY'S SATIONARY OFFICE (HMSO) 2004. Human Tissue Act 2004. HMSO London: Her Majesty's Stationary Office.

HER MAJESTY'S SATIONARY OFFICE (HMSO) 2012. Health and Social Care Act 2012. Her Majesty's Stationary Office (HMSO).

HER MAJESTY'S SATIONARY OFFICE (HSMO) 2001. Health and Social Care Act 2001. HMSO London: Her Majesty's Stationary Office (HMSO).

HER MAJESTY'S STATIONARY OFFICE 1998. Data Protection Act 1998. *In:* LEGISLATION.GOV.UK (ed.). London.

HER MAJESTY'S STATIONARY OFFICE (HMSO) 2006. National Health Service Act 2006 HMSO London: Her Majesty's Stationary Office (HMSO).

HILLESTAD, R., BIGELOW, J., BOWER, A., GIROSI, F., MEILI, R., SCOVILLE, R. & TAYLOR, R. 2005. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Aff (Millwood),* 24**,** 1103-17.

HORNER, J. S. 1998. Research, ethics and privacy: the limits of knowledge. *Public Health,* 112**,** 217-20.

HUGHES, S., WELLS, K., MCSORLEY, P. & FREEMAN, A. 2014. Preparing individual patient data from clinical trials for sharing: the GlaxoSmithKline approach. *Pharm Stat*.

HUMPHREYS, S., THOMAS, H. & MARTIN, R. 2014. Medical Dominance within Research Ethics Committees. *Account Res,* 21**,** 366-88.

HUNTINGTON, S., THORNE, C., ANDERSON, J., NEWELL, M. L., TAYLOR, G. P., PILLAY, D., HILL, T., TOOKEY, P. & SABIN, C. 2014. Response to antiretroviral therapy (ART): comparing women with previous use of zidovudine monotherapy (ZDVm) in pregnancy with ART naive women. *BMC Infect Dis,* 14**,** 127.

HUNTINGTON, S. E., BANSI, L. K., THORNE, C., ANDERSON, J., NEWELL, M. L., TAYLOR, G. P., PILLAY, D., HILL, T., TOOKEY, P. A. & SABIN, C. A. 2012. Using two on-going HIV studies to obtain clinical data from before, during and after pregnancy for HIV-positive women. *BMC Med Res Methodol,* 12**,** 110.

I-BASE. 2014. *i-base Website* [Online]. Available: http://i-base.info/ [Accessed 27th October 2014].

IBM. 2003. *Enterprise Privacy Authorization Language (EPAL)* [Online]. Available: http://www.zurich.ibm.com/security/enterprise-privacy/epal/Specification/index.html [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2009. *Privacy Impact Assessments (PIA)* [Online]. Information Commissioner's Office. Available: http://www.ico.gov.uk/for_organisations/data_protection/topic_guides/privacy_impact_assessment.aspx [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2011a. *Data Sharing Code of Practice* [Online]. Information Commissioner's Office. Available: http://www.ico.gov.uk/~/media/documents/library/Data_Protection/Detailed_specialist_guides/data_sharing_code_of_practice.pdf [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2011b. *Information Commissioner's Office Website: Data Protection Technical Guidance, Determining what is personal data* [Online]. ICO. Available: http://ico.org.uk/~/media/documents/library/Data_Protection/Detailed_specialist_guides/PERSONAL_DATA_FLOWCHART_V1_WITH_PREFACE001.ashx [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2011c. *Information Commissioner's Office Website: Determining what information is 'data' for the purposes of the DPA* [Online]. Available: http://ico.org.uk/~/media/documents/library/data_protection/detailed_specialist_guides/what_is_data_for_the_purposes_of_the_dpa.pdf [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2014a. *Information Commissioner's Office* [Online]. Information Commissioner's Office. Available: http://www.ico.gov.uk/ [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2014b. *Information Commissioner's Office Website* [Online]. ICO. Available: http://www.ico.gov.uk/ [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2014c. *Information Commissioner's Office Website: Subject Access Code of Practice, Dealing with requests from individuals for personal information* [Online]. Available: http://ico.org.uk/for_organisations/data_protection/~/media/documents/library/Data_Protection/Detailed_specialist_guides/subject-access-code-of-practice.PDF [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2014d. *Information Commissioner's Office: Key Definitions of the Data Protection Act* [Online]. Available:

http://ico.org.uk/for_organisations/data_protection/the_guide/key_definitions [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2014e. *Information Commissioner's Office Website: Anonymisation* [Online]. Available: http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation [Accessed 27th October 2014].

INFORMATION COMMISSIONER'S OFFICE. 2014f. *The Information Commissioner's Office Website: Enforcement* [Online]. Available: http://ico.org.uk/enforcement [Accessed 27th October 2014].

INFORMATION TECHNOLOGY GOVERNANCE INSTITUTE. 2014. *IT Governance Institute Website* [Online]. ITGI. Available: http://www.itgi.org/ [Accessed 27th October 2014].

INNOVATIVE MEDICINES INITIATIVE. 2014. *The Innovative Medicines Initiative Website* [Online]. Innovative Medicines Initiative. Available: http://www.imi.europa.eu/ [Accessed 27th October 2014].

INSIDE GOVERNMENT. 2013. *The Changing Landscape of Information Governance in Health and Social Care Services Conference Website* [Online]. Available: http://www.insidegovernment.co.uk/event-details/information-governance/148/ [Accessed 27th October 2014].

INTERNATIONAL HEALTH TERMINOLOGY STANDARDS DEVELOPMENT ORGANISATION (IHTSDO). 2014. *SNOMED CT Website* [Online]. International Health Terminology Standards Development Organisation (IHTSDO). Available: http://www.ihtsdo.org/snomed-ct/ [Accessed 27th October 2014].

INTERNATIONAL ORGANISATION FOR STANDARDIZATION (ISO) 2011. ISO/TS 14265:2011 Health Informatics - Classification of purposes for processing personal health information.

INTERNATIONAL ORGANISATION FOR STANDARDIZATION (ISO). 2014. *International Organization for Standardization Website* [Online]. ISO. Available: http://www.iso.org/iso/home.html [Accessed 27th October 2014].

JACOBSON, I., BROOCH, G. & RAMBAUGH, J. 1999. *The Unified Software Development Process*, Addison Wesley.

JAMAL, L., SAPP, J. C., LEWIS, K., YANES, T., FACIO, F. M., BIESECKER, L. G. & BIESECKER, B. B. 2013. Research participants' attitudes towards the confidentiality of genomic sequence information. *Eur J Hum Genet*.

JAMES R. LEWIS 1993. IBM Computer Usability Satisfaction Questionnaires Psycho Metric Evaluation and Instructions for Use. IBM.

JONES, K. H., FORD, D. V., JONES, C., DSILVA, R., THOMPSON, S., BROOKS, C. J., HEAVEN, M. L., THAYER, D. S., MCNERNEY, C. L. & LYONS, R. A. A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation. *Journal of Biomedical Informatics,* 50**,** 196-204.

JUNGHANS, C., FEDER, G., HEMINGWAY, H., TIMMIS, A. & JONES, M. 2005. Recruiting patients to medical research: double blind randomised trial of "opt-in" versus "opt-out" strategies. *BMJ,* 331**,** 940.

KALRA, D. 2002. *Clinical Foundations and Information Achitecture for the Implementation of a Federated Health Record Service.* PhD PhD, UCL.

KALRA, D. 2006. Electronic health record standards. *Yearb Med Inform***,** 136-44.

KALRA, D. & FERNANDO, B. 2013. A review of the empirical evidence of the healthcare benefits of personal health records. *Yearb Med Inform,* 8**,** 93-102.

KALRA, D. & INGRAM, D. 2006. Ethical issues of electronic patient data and informatics in clinical trial settings. *In:* NAGL, S. (ed.) *Cancer bioinformatics: from therapy to treatment.* John Wiley & Sons.

KALRA, D., SINGLETON, P., MILAN, J., MACKAY, J., DETMER, D., RECTOR, A. & INGRAM, D. 2005. Security and confidentiality approach for the Clinical E-Science Framework (CLEF). *Methods Inf Med,* 44**,** 193-7.

KING, T., BRANKOVIC, L. & GILLARD, P. 2012. Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. *Int J Med Inform,* 81**,** 279-89.

KING'S HEALTH PARTNERS. 2014. *Cancer Biobank Website* [Online]. Available: http://www.kingshealthpartners.org/info/cancer-biobank [Accessed 27th October 2014].

KOBAYASHI, S., KIMURA, E. & ISHIHARA, K. 2013. Archetype Model-Driven Development Framework for EHR Web System. *Healthc Inform Res,* 19**,** 271-7.

KUZU, M., KANTARCIOGLU, M., DURHAM, E. & MALIN, B. 2011. A Constraint Satisfaction Cryptanalysis of Bloom Filters in Private Record Linkage. *In:* FISCHER-HÜBNER, S. & HOPPER, N. (eds.) *Privacy Enhancing Technologies.* Springer Berlin Heidelberg.

KUZU, M., KANTARCIOGLU, M., INAN, A., BERTINO, E., DURHAM, E. & MALIN, B. 2013. Efficient Privacy-Aware Record Integration. *Adv Database Technol***,** 167-178.

LANGANKE, M., BROTHERS, K. B., ERDMANN, P., WEINERT, J., KRAFCZYK-KORTH, J., DORR, M., HOFFMANN, W., KROEMER, H. K. & ASSEL, H. 2011. Comparing different scientific approaches to personalized medicine: research ethics and privacy protection. *Per Med,* 8**,** 437-44.

LAURIE, G. & SETHI, N. 2013. Towards Principles-Based Approaches to Governance of Health-related Research using Personal Data. *Eur J Risk Regul,* 4**,** 43-57.

LEA, N., AUSTIN, T., HAILES, S. & KALRA, D. 2009. Expression of Security Policy in Medical Systems for Electronic Healthcare Records. Worl Academy of Science, Engineering and Technology 2009, May 2009 2009 Tokyo. 711-715.

LEA, N., HAILES, S., AUSTIN, T. & KALRA, D. 2008. Knowledge management for the protection of information in electronic medical records. *Stud Health Technol Inform,* 136**,** 685-90.

LESLIE, H. 2012. *openEHR Website: Introduction to Archetypes and Archetype Classes* [Online]. openEHR Foundation. Available: http://www.openehr.org/wiki/display/healthmod/Introduction+to+Archetypes+and+Archetype+classes [Accessed 27th October 2014].

LEZCANO, L., SICILIA, M. A. & RODRIGUEZ-SOLANO, C. 2011. Integrating reasoning and clinical archetypes using OWL ontologies and SWRL rules. *J Biomed Inform,* 44**,** 343-53.

LI, F., ZOU, X., LIU, P. & CHEN, J. Y. 2011. New threats to health data privacy. *BMC Bioinformatics,* 12 Suppl 12**,** S7.

LIANG, S. F., TAWEEL, A., MILES, S., KOVALCHUK, Y., SPIRIDOU, A., BARRATT, B., HOANG, U., CRICHTON, S., DELANEY, B. C. & WOLFE, C. 2014. Semi Automated Transformation to OWL Formatted Files as an Approach to Data Integration. A Feasibility Study Using Environmental, Disease Register and Primary Care Clinical Data. *Methods Inf Med,* 53.

LIN, A. & BROWN, R. 2000. The application of security policy to role-based access control and the common data security architecture. *Computer Communications,* 23**,** 1584-1593.

LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE. 2014. *ME-CFS BioBank Website* [Online]. Available: http://www.lshtm.ac.uk/itd/crd/research/cure-me/ukmecfsbiobank/ [Accessed 27th October 2014].

LOUKIDES, G. & GKOULALAS-DIVANIS, A. 2012. Utility-preserving transaction data anonymization with low information loss. *Expert Syst Appl,* 39**,** 9764-9777.

LYONS, R. A., HUTCHINGS, H., RODGERS, S. E., HYATT, M. A., DEMMLER, J., GABBE, B. J., BROOKS, C. J., BROPHY, S., JONES, K., FORD, D. V., PARANJOTHY, S., FONE, D., DUNSTAN, F. D., EVANS, A., KELLY, M., WATKINS, W. J., MADDOCKS, A., BARNES, P., JAMES-ELLISON, M., JOHN, G. & LOWE, S. 2012. Development and use of a privacy-protecting total population record linkage system to support observational, interventional, and policy relevant research. *The Lancet,* 380**,** S6.

MALDONADO, J. A., COSTA, C. M., MONER, D., MENARGUEZ-TORTOSA, M., BOSCA, D., MINARRO GIMENEZ, J. A., FERNANDEZ-BREIS, J. T. & ROBLES, M. 2012.

Using the ResearchEHR platform to facilitate the practical application of the EHR standards. *J Biomed Inform,* 45**,** 746-62.

MALDONADO, J. A., MONER, D., BOSCA, D., FERNANDEZ-BREIS, J. T., ANGULO, C. & ROBLES, M. 2009. LinkEHR-Ed: a multi-reference model archetype editor based on formal semantics. *Int J Med Inform,* 78**,** 559-70.

MANION, F. J., ROBBINS, R. J., WEEMS, W. A. & CROWLEY, R. S. 2009. Security and privacy requirements for a multi-institutional cancer research data grid: an interview-based study. *BMC Med Inform Decis Mak,* 9**,** 31.

MARTINEZ, S., SANCHEZ, D. & VALLS, A. 2013. A semantic framework to protect the privacy of electronic health records with non-numerical attributes. *J Biomed Inform,* 46**,** 294-303.

MARTINEZ-COSTA, C., MENARGUEZ-TORTOSA, M. & FERNANDEZ-BREIS, J. T. 2010. An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. *J Biomed Inform,* 43**,** 736-46.

MARTINEZ-COSTA, C., MENARGUEZ-TORTOSA, M., FERNANDEZ-BREIS, J. T. & MALDONADO, J. A. 2009. A model-driven approach for representing clinical archetypes for Semantic Web environments. *J Biomed Inform,* 42**,** 150-64.

MAY, M., GOMPELS, M., DELPECH, V., PORTER, K., POST, F., JOHNSON, M., DUNN, D., PALFREEMAN, A., GILSON, R., GAZZARD, B., HILL, T., WALSH, J., FISHER, M., ORKIN, C., AINSWORTH, J., BANSI, L., PHILLIPS, A., LEEN, C., NELSON, M., ANDERSON, J. & SABIN, C. 2011. Impact of late diagnosis and treatment on life expectancy in people with HIV-1: UK Collaborative HIV Cohort (UK CHIC) Study. *Bmj,* 343**,** d6016.

MEDICAL RESEARCH COUNCIL 2012. MRC Ethics Series - Good Research Practice: Principles and Guidelines. Medical Research Council.

MEDICAL RESEARCH COUNCIL. 2014a. *Data Management Plans Website* [Online]. Medical Research Council. Available: http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/data-management-plans/ [Accessed 27th October 2014].

MEDICAL RESEARCH COUNCIL. 2014b. *Medical Research Council Website: Initiatives in Informatics Research* [Online]. Medical Research Council. Available: http://www.mrc.ac.uk/research/initiatives/health-and-biomedical-informatics/initiatives-in-informatics-research/ [Accessed 27th October 2014].

MEDICAL RESEARCH COUNCIL. 2014c. *UK Register of HIV Seroconverters Website* [Online]. Available: http://www.ctu.mrc.ac.uk/our_research/research_areas/hiv/studies/ukr/ [Accessed 27th October 2014].

MEIZOSO GARCIA, M., IGLESIAS ALLONES, J. L., MARTINEZ HERNANDEZ, D. & TABOADA IGLESIAS, M. J. 2012. Semantic similarity-based alignment between clinical archetypes and SNOMED CT: an application to observations. *Int J Med Inform,* 81**,** 566-78.

MINARRO-GIMENEZ, J. A., BLAGEC, K., BOYCE, R. D., ADLASSNIG, K. P. & SAMWALD, M. 2014. An ontology-based, mobile-optimized system for pharmacogenomic decision support at the point-of-care. *PLoS One,* 9**,** e93769.

MOHAMMED, N., JIANG, X., CHEN, R., FUNG, B. C. & OHNO-MACHADO, L. 2013. Privacy-preserving heterogeneous health data sharing. *J Am Med Inform Assoc,* 20**,** 462-9.

MONER, D., MALDONADO, J. A., ANGULO, C., BOSCA, D., PEREZ, D., ABAD, I., REIG, E. & ROBLES, M. 2008. Standardization of discharge reports with the ISO 13606 norm. *Conf Proc IEEE Eng Med Biol Soc,* 2008**,** 1470-3.

MOTTA, G. H. & FURUIE, S. S. 2003. A contextual role-based access control authorization model for electronic patient record. *IEEE Trans Inf Technol Biomed,* 7**,** 202-7.

NATIONAL AUDIT OFFICE 2011. The National Programme for IT in the NHS: an update on the delivery of detailed care records systems.

NATIONAL HEALTH SERVICE. 2011a. *The National Research Ethics Service Website* [Online]. Available: http://www.nres.nhs.uk/ [Accessed 27th October 2014].

NATIONAL HEALTH SERVICE. 2011b. *NRES Website: The Role of Research Ethics Committees* [Online]. Available: http://www.nres.nhs.uk/about-the-national-research-ethics-service/about-recs/role-of-recs/ [Accessed 27th October 2014].

NATIONAL HEALTH SERVICE. 2014a. *The Health Research Athority Website: Confidentiality Advisory Group* [Online]. NHS. Available: http://www.hra.nhs.uk/resources/confidentiality-advisory-group/ [Accessed 27th October 2014].

NATIONAL HEALTH SERVICE. 2014b. *The Health Research Authority Website* [Online]. Available: http://www.hra.nhs.uk/ [Accessed 27th October 2014].

NATIONAL HEALTH SERVICE. 2014c. *The Health Research Authority Website: Intergrated Research Application System (IRAS)* [Online]. UK: NHS. Available: http://www.hra.nhs.uk/resources/applying-for-reviews/integrated-research-application-system-iras/ [Accessed 27th October 2014].

NATIONAL HEALTH SERVICE. 2014d. *The Health Research Authority Website: Research Community* [Online]. NHS. Available: http://www.hra.nhs.uk/research-community/ [Accessed 27th October 2014].

NATIONAL HEALTH SERVICE. 2014e. *Research Ethics Committees (RECs)* [Online]. NHS. Available: http://www.hra.nhs.uk/about-the-hra/our-committees/research-ethics-committees-recs/ [Accessed 27th October 2014].

NATIONAL INFORMATION GOVERNANCE BOARD (NIGB). 2013a. *The National Information Governance Board (NIGB) Section 251 Exemption* [Online]. Available: http://webarchive.nationalarchives.gov.uk/20130513181011/http://www.nigb.nhs.uk/s251 [Accessed 27th October 2014].

NATIONAL INFORMATION GOVERNANCE BOARD (NIGB). 2013b. *National Information Governance Board Website* [Online]. Available: http://webarchive.nationalarchives.gov.uk/20130513181011/http://www.nigb.nhs.uk/ [Accessed 27th October 2014].

NATIONAL INFORMATION GOVERNANCE BOARD (NIGB). 2013c. *Teh NIGB Website Ethics & Confidentiality Committee* [Online]. Available: http://webarchive.nationalarchives.gov.uk/20130513181011/http://www.nigb.nhs.uk/ecc [Accessed 27th October 2014].

NATIONAL INSTITUTE FOR HEALTH RESEARCH. 2013. *NIHR Website: Major Health Informatics Programme Announced* [Online]. National Institute for Health Research. Available: http://www.uclhospitals.brc.nihr.ac.uk/news/major-health-informatics-programme-announced [Accessed 27th October 2014].

NATIONAL INSTITUTE FOR HEALTH RESEARCH. 2014a. *Biomedical Research Centre Oxford Website* [Online]. National Institute for Health Research. Available: http://oxfordbrc.nihr.ac.uk/ [Accessed 27th October 2014].

NATIONAL INSTITUTE FOR HEALTH RESEARCH. 2014b. *National Institute for Health Research BioResource Website* [Online]. Available: http://bioresource.nihr.ac.uk [Accessed 27th October 2014].

NATURE EDITORIAL 2014a. Careless.data. *Nature,* 507**,** 7.

NATURE EDITORIAL 2014b. Power to the People. *Nature,* 505.

NEUBAUER, T. & HEURIX, J. 2011. A methodology for the pseudonymization of medical data. *Int J Med Inform,* 80**,** 190-204.

NOUMEIR, R., LEMAY, A. & LINA, J. M. 2007. Pseudonymization of radiology data for research purposes. *J Digit Imaging,* 20**,** 284-95.

NUNNALLY, J. C. 1978. *Psychometric Theory* McGraw Hill.

NYREN, O., STENBECK, M. & GRONBERG, H. 2014. The European Parliament proposal for the new EU General Data Protection Regulation may severely restrict European epidemiological research. *Eur J Epidemiol,* 29**,** 227-30.

O'NEILL, O. 2003. Some limits of informed consent. *Journal of Medical Ethics,* 29**,** 4-7.

ODERSKY, M., SPOON, L. & VENNERS, B. 2010. *Programming in Scala,* USA, Artima Press.

OFFICE OF POPULATION CENSUSES AND SURVEYS (OPCS). 2014. *HSCIC Website: OPCS Classification of Intervention and Procedures* [Online]. Office of Population Censuses and Surveys (OPCS). Available: http://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/opcs_classification_of_interventions_and_procedures.asp?shownav=1 [Accessed 27th October 2014].

OHNO-MACHADO, L., VINTERBO, S. A. & DREISEITL, S. 2001. Effects of data anonymization by cell suppression on descriptive statistics and predictive modeling performance. *Proc AMIA Symp***,** 503-7.

OPENEHR 2014a. Knowledge Artefact Identification.

OPENEHR. 2014b. *The openEHR Foundation Website* [Online]. Available: http://www.openehr.org [Accessed 26/06/2014 2014].

OPENEHR. 2014c. *openEHR Website: About ADL 1.5* [Online]. openEHR. Available: http://www.openehr.org/downloads/ADLworkbench/learning_about [Accessed 27th October 2014].

OPENEHR. 2014d. *openEHR Website: ADL Workbench* [Online]. openEHR. Available: http://www.openehr.org/downloads/ADLworkbench/home [Accessed 27th October 2014].

OPENEHR. 2014e. *openEHR Website: Clinical Knowledge Manager* [Online]. openEHR. Available: http://www.openehr.org/ckm/ [Accessed 27th October 2014].

OPENEHR. 2014f. *openEHR Website: Obtaining Archetypes* [Online]. openEHR. Available: http://www.openehr.org/downloads/ADLworkbench/obtaining_archetypes [Accessed 27th October 2014].

OPENEHR. 2014g. *openEHR Website: Vendor / Developer Guide to openEHR* [Online]. openEHR. Available: http://www.openehr.org/getting_involved/vendors_developers [Accessed 27th October 2014].

OPENEHR. 2014h. *openEHR Website: What is openEHR?* [Online]. openEHR. Available: http://www.openehr.org/what_is_openehr [Accessed 27th October 2014].

ORGANIZATION FOR THE ADVANCEMENT OF STRUCTURED INFORMATION STANDARDS. 2011a. *OASIS Website: Extensible Access Control Markup Language (XACML)* [Online]. Available: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml [Accessed 27th October 2014].

ORGANIZATION FOR THE ADVANCEMENT OF STRUCTURED INFORMATION STANDARDS. 2011b. *OASIS Website: Security Assertion Markup Language (SAML)* [Online]. OASIS Foundation. Available: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security [Accessed 27th October 2014].

ORGANIZATION FOR THE ADVANCEMENT OF STRUCTURED INFORMATION STANDARDS. 2011c. *Organization for the Advancement of Structured Information Standards Website* [Online]. Available: http://www.oasis-open.org/ [Accessed 27th October 2014].

OXFORD BIOBANK. 2012. *Oxford BioBank Website* [Online]. Available: http://www.oxfordbiobank.org.uk [Accessed 27th October 2014].

OXFORD INTERNET INSTITUTE. 2005. *Oxford Internet Institite Website: Ethics in e-Science* [Online]. The Oxford Internet Institute. Available: http://www.oii.ox.ac.uk/events/?id=54 [Accessed 27th October 2014].

OXFORD UNIVERSITY. 2002. *Oxford University Website: eDiaMoND: What is eDiamond?* [Online]. Available: http://www.ediamond.ox.ac.uk/whatis.html [Accessed 19/09/2014 2014].

OXFORD UNIVERSITY. 2008. *Oxford Univesity Department of Computer Science Website: CancerGrid: A consortium to develop open standards for clinical cancer informatics* [Online]. Available: http://www.cs.ox.ac.uk/projects/cancergrid/ [Accessed 27th October 2014].

P-MEDICINE. 2014. *p-medicine Website: p-medicine in brief* [Online]. p-medicine. Available: http://p-medicine.eu/project/in-brief/ [Accessed 27th October 2014].

PARTRIDGE, S. N. 2014. Review of data releases by the NHS Information Centre.

PERERA, G., HOLBROOK, A., THABANE, L., FOSTER, G. & WILLISON, D. J. 2011. Views on health information sharing and privacy from primary care practices using electronic medical records. *Int J Med Inform,* 80**,** 94-101.

PRICEWATERHOUSECOOPERS LLP 2013. A review of the potential benefits from the better use of information and technology in Health and Social Care: Final report. United Kingdom.

PUBLIC HEALTH ENGLAND. 2014. *Public Health England Website: HIV: surveillance, data and management* [Online]. Available: https://www.gov.uk/government/collections/hiv-surveillance-data-and-management [Accessed 27th October 2014].

QAMAR, R. & RECTOR, A. 2007. Semantic Mapping of Clinical Model Data to Biomedical Terminologies to Facilitate Data Interoperability. *Healthcare Computing 2007 Conference.* Harrogate, UK.

RESEARCH, N. I. F. H. 2014. *NIHR Website: Biomedical Research Centre UCLH* [Online]. National Institute for Health Research. Available: http://www.uclhospitals.brc.nihr.ac.uk/ [Accessed 27th October 2014].

RINNER, C., JANZEK-HAWLAT, S., SIBINOVIC, S. & DUFTSCHMID, G. 2010. Semantic validation of standard-based electronic health record documents with W3C XML schema. *Methods Inf Med,* 49**,** 271-80.

ROGERS, J., TAWEEL, A., RECTOR, A., KALRA, D., INGRAM, D., MILAN, J., SINGLETON, P., GAIZAUSKAS, R., HEPPLE, M., SCOTT, D. & POWER, R. 2004. A Co-operative Clinical E-Science Framework

(CLEF): Joining up Healthcare and Clinical Research. *eScience All Hands Meeting 2004.* Nottingham: JISC.

ROOM, S. 2004. Data protection, informed consent, and research: Data Protection Act does not bar medical research. *British Medical Journal (BMJ),* 328**,** 1437.

ROYAL COLLEGE OF GENERAL PRACTITIONERS 2013. RCGP Policy statement Information Governance of the Use of Personal Confidential Data Held by General Practices for Secondary Purposes.

ROYAL COLLEGE OF GENERAL PRACTITIONERS. 2014. *RCGP Website: RCGP Voices Concern over Care.data* [Online]. Available: http://www.rcgp.org.uk/news/2014/february/rcgp-voices-concerns-about-care-data.aspx [Accessed 27th October 2014].

ROYAL COLLEGE OF PRACTITIONERS. 2004. *Royal College of Physicians Symposium on the Governance of Medical Research Databases* [Online]. [Accessed 27th October 2014].

ROYAL STATISTICAL SOCIETY 2014. Royal Statistical Society research on trust in data and attitudes toward data use / data sharing.

RYNNING, E. 2007. Public trust and privacy in shared electronic health records. *Eur J Health Law,* 14**,** 105-12.

SANCHEZ-DE-MADARIAGA, R., MUNOZ, A., CACERES, J., SOMOLINOS, R., PASCUAL, M., MARTINEZ, I., SALVADOR, C. H. & MONTEAGUDO, J. L. 2013. ccML, a new mark-up language to improve ISO/EN 13606-based electronic health record extracts practical edition. *J Am Med Inform Assoc,* 20**,** 298-304.

SANDHU, R. S., COYNE, E. J., FEINSTEIN, H. L. & YOUMAN, C. E. 1996. Role-Based Access Control Models. *Computer,* 29**,** 38-47.

SANDHU, R. S. & SAMARATI, P. 1994. Access control: principle and practice. *Communications Magazine, IEEE,* 32**,** 40-48.

SANDS, D. Z. & WALD, J. S. 2014. Transforming health care delivery through consumer engagement, health data transparency, and patient-generated health information. *Yearb Med Inform,* 9**,** 170-6.

SANTOS, M. R., BAX, M. P. & KALRA, D. 2010. Building a logical EHR architecture based on ISO 13606 standard and semantic web technologies. *Stud Health Technol Inform,* 160**,** 161-5.

SAVULESCU, J., CHALMERS, I. & BLUNT, J. 1996. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *Bmj,* 313**,** 1390-3.

SCHNELL, R., BACHTELER, T. & REIHER, J. 2009. Privacy-preserving record linkage using Bloom filters. *BMC Medical Informatics and Decision Making,* 9**,** 41.

SCHUTZE, B. 2013. [Legal framework of data protection : current requirements in Germany and requirements in planned European Union regulations]. *Radiologe,* 53**,** 437-40.

SCOTTISH HEALTH INFORMATICS PROGRAMME. 2014. *SHIP Website: The Scottish Health Informatics Programme* [Online]. Scotland. Available: http://www.scot-ship.ac.uk/ [Accessed 27th October 2014].

SETHI, N. & LAURIE, G. T. 2013. Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together. *Med Law Int,* 13**,** 168-204.

SEVENTH FRAMEWORK PROGRAMME. 2014. *DebugIT Website* [Online]. Available: http://www.debugit.eu [Accessed 27th October 2014].

SHEATHER, J. & BRANNAN, S. 2013. Patient confidentiality in a time of care.data. *Bmj,* 347**,** f7042.

SHERWOOD, J., CLARK, A. & LYNAS, D. 2005. *Enterprise Security Architecture - A Business Driven Approach*, CMP Books.

SLOMAN, M. & LUPU, E. 2002. Security and Management Policy Specification. *IEEE Network Special issue on Policy Based Networking,* 16**,** 10-19.

SLOWTHER, A., BOYNTON, P. & SHAW, S. 2006. Research Governance: Ethical Issues. *Journal of the Royal Society of Medicine,* 99**,** 65-72.

SMITH, D. & ELLIOT, M. 2008. A Measure of Disclosure Risk for Tables of Counts. *Transactions on Data Privacy,* 1**,** 34.

STANDARDIZATION, T. E. C. F. 2014. *The European Committee for Standardization Website: Who We Are* [Online]. Available: https://www.cen.eu/about/Pages/default.aspx [Accessed 27th October 2014].

STANDARDS, O. F. T. A. O. S. I. 2011. *OASIS Website: Privacy Management Reference Model Technical Committee* [Online]. OASIS. Available: http://www.oasis-open.org/committees/pmrm/charter.php [Accessed 27th October 2014].

STEVENSON, F., LLOYD, N., HARRINGTON, L. & WALLACE, P. 2013. Use of electronic patient records for research: views of patients and staff in general practice. *Fam Pract,* 30**,** 227-32.

SUN, S., AUSTIN, T. & KALRA, D. 2012. A Data Types Profile Suitable for Use with ISO EN 13606. *J. Med. Syst.,* 36**,** 3621-3635.

SUNDVALL, E., QAMAR, R., NYSTROM, M., FORSS, M., PETERSSON, H., KARLSSON, D., AHLFELDT, H. & RECTOR, A. 2008. Integration of tools for binding archetypes to SNOMED CT. *BMC Med Inform Decis Mak,* 8 Suppl 1**,** S7.

SWANSEA UNIVERSITY. 2013. *The Secure Anonymised Information Linkage Databank Website* [Online]. Swansea. Available: http://www.saildatabank.com [Accessed 27th October 2014].

SWEENEY, L. 2002. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems,* 10**,** 557-570.

TAMERSOY, A., LOUKIDES, G., NERGIZ, M. E., SAYGIN, Y. & MALIN, B. 2012. Anonymization of longitudinal electronic medical records. *IEEE Trans Inf Technol Biomed,* 16**,** 413-23.

TAPURIA, A., KALRA, D. & KOBAYASHI, S. 2013. Contribution of Clinical Archetypes, and the Challenges, towards Achieving Semantic Interoperability for EHRs. *Healthc Inform Res,* 19**,** 286-92.

THE CALDICOTT COMMITTEE 1997. Report on the Review of Patient-Identifiable Information. Department of Health.

THE ESCIENCE INITIATIVE. 2014. *The National eScience Centre Website,* [Online]. Available: http://www.nesc.ac.uk/ [Accessed 27th October 2014].

THE HEALTH AND SOCIAL CARE INFORMATION CENTRE 2013. A guide to confidentiality in health and social care.

THE INTERNATIONAL ORGANISATION FOR STANDARDIZATION 2008. ISO 27799: 2008 Health informatics -- Information security management in health using ISO/IEC 27002. International Organization for Standardization.

THE INTERNATIONAL ORGANISATION FOR STANDARDIZATION. 2011. *The International Organisation for Standardisation (ISO)* [Online]. Available: http://www.iso.org/ [Accessed 06/04/2011 2011].

THE NHS FUTURE FORUM 2012. Information - A Report from teh NHS Future Forum. UK: NHS.

THOMAS, R. & WALPORT, M. 2008. Data Sharing Review.

THOMSON, C. 2012. Research Ethics Committees. *In:* CHADWICK, R. (ed.) *Encyclopedia of Applied Ethics (Second Edition).* San Diego: Academic Press.

TIMES NEWSPAPER. 2006. Medical Research Need not Fall Foul of the Law. *The Times Newspaper*, 20/01/2006.

TOCCACELI, V. & MASOCCO, M. 2012. [The use of sensitive data in epidemiology: remarks on the difficulties when interpreting the Italian legislation]. *Epidemiol Prev,* 36**,** 280-6.

TRANSFORM PROJECT. 2014. *The TRANSFoRm Project Website* [Online]. Available: http://www.transformproject.eu [Accessed 27th October 2014].

TREVENA, L., IRWIG, L. & BARRATT, A. 2006. Impact of privacy legislation on the number and characteristics of people who are recruited for research: a randomised trial. *Journal of Medical Ethics,* 32**,** 473 - 477.

TRUST, S. L. A. M. N. 2014. *SLAM Website: Bimoedical Research Centre South London and Maudsley* [Online]. National Institute for Health Research. Available: http://brc.slam.nhs.uk/ [Accessed 27th October 2014].

UK BIOBANK. 2011. *UK Biobank Website: About UK Biobank* [Online]. Available: http://www.ukbiobank.ac.uk/about-biobank-uk/ [Accessed 17/08/2014 2014].

UK COLLABORATIVE GROUP ON HIV DRUG RESISTANCE, UK COLLABORATIVE HIV COHORT STUDY & UK REGISTER OF HIV SEROCONVERTERS 2007. Evidence of a decline in transmitted HIV-1 drug resistance in the United Kingdom. *Aids,* 21**,** 1035-9.

UK COLLABORATIVE HIV COHORT STUDY. 2014. *The UK Collaborative HIV Cohort Study Website* [Online]. Available: http://www.ctu.mrc.ac.uk/UKCHIC/indexUKCHIC.asp [Accessed 27th October 2014].

UK COMMUNITY ADVISORY BOARD. 2014. *The UK Community Advisory Board Website* [Online]. Available: http://www.ukcab.net/ [Accessed 27th October 2014].

UK HIV DRUG RESISTANCE DATABASE. 2014. *UK HIV Drug Resistance Database Website* [Online]. Available: http://www.ctu.mrc.ac.uk/hivrdb/public/default.asp [Accessed 27th October 2014].

UNIVERSITY COLLEGE LONDON. 2010a. *UCL Website: Research Governace* [Online]. Available: http://www.ucl.ac.uk/srs/governance-and-committees/resgov [Accessed 27th October 2014].

UNIVERSITY COLLEGE LONDON. 2010b. *UCL Website: The Databases for HIV: Integration, Collaboration and Engagement Initiative* [Online]. Available: http://www.ucl.ac.uk/iph/research/hivbiostatistics/dhice [Accessed 27th October 2014].

UNIVERSITY COLLEGE LONDON. 2013. *UCL Website: The National Study of HIV in Pregnance and Childhood* [Online]. Available: http://www.ucl.ac.uk/nshpc [Accessed 27th October 2014].

UNIVERSITY COLLEGE LONDON. 2014a. *UCL Website: Biobanking at UCL* [Online]. Available: http://www.ucl.ac.uk/biobank [Accessed 27th October 2014].

UNIVERSITY COLLEGE LONDON. 2014b. *UCL Website: Exemptions* [Online]. UCL. Available: http://ethics.grad.ucl.ac.uk/exemptions.php [Accessed 27th October 2014].

UNIVERSITY COLLEGE LONDON. 2014c. *UCL Website: Which Research Ethics Committee Do I Apply to?* [Online]. UCL. Available: http://ethics.grad.ucl.ac.uk/which-ethics-committee-apply-to.php [Accessed 27th October 2014].

US NATIONAL LIBRARY OF MEDICINE. 2014. *Unified Medical Langauge System (UMLS) Website* [Online]. US National Library of Medicine. Available: http://www.nlm.nih.gov/research/umls/ [Accessed 27th October 2014].

VAN STAA, T. P., DYSON, L., MCCANN, G., PADMANABHAN, S., BELATRI, R., GOLDACRE, B., CASSELL, J., PIRMOHAMED, M., TORGERSON, D., RONALDSON, S., ADAMSON, J., TAWEEL, A., DELANEY, B., MAHMOOD, S., BARACAIA, S., ROUND, T., FOX, R., HUNTER, T., GULLIFORD, M. & SMEETH, L. 2014. The opportunities and challenges of pragmatic point-of-care

randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess,* 18.

VIRONE, M. G. 2012. EHR and data protection issues in Italy. *Stud Health Technol Inform,* 180**,** 741-5.

WALLEY, T. 2006. Using personal health information in medical research. *British Medical Journal (BMJ),* 332**,** 130-131.

WARD, H. J. T., COUSENS, S. N., SMITH-BATHGATE, B., LEITCH, M., EVERINGTON, D., WILL, R. G. & SMITH, P. G. 2004. Obstacles to conducting epidemiological research in the UK general population. *BMJ,* 329**,** 277-279.

WATSON, N. 2006. Patients should have to opt out of national electronic care recordsFORAGAINST. *BMJ,* 333**,** 39-42.

WEED, L. 1968. Problem-based patient record. *N Engl J Med,* 278.

WELLCOME TRUST 2014a. Protecting health and scientific research in the Data Protection Regulation (2012/0011(COD))

Position of non-commercial research organisations and academics – July 2014. The Wellcome Trust.

WELLCOME TRUST. 2014b. *Wellcome Trust Website: Guidance for researchers: Developing a data management and sharing plan* [Online]. The Wellcome Trust. Available: http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/ [Accessed 27th October 2014].

WHIDDETT, R., HUNTER, I., ENGELBRECHT, J. & HANDY, J. 2006. Patients' attitudes towards sharing their health information. *Int J Med Inform,* 75**,** 530-41.

WIESENAUER, M., JOHNER, C. & ROHRIG, R. 2012. Secondary use of clinical data in healthcare providers - an overview on research, regulatory and ethical requirements. *Stud Health Technol Inform,* 180**,** 614-8.

WILLIAMS, B., DOWELL, J., HUMPHRIS, G., THEMESSL-HUBER, M., RUSHMER, R., RICKETTS, I., BOYLE, P. & SULLIVAN, F. 2010. Developing a longitudinal database of routinely recorded primary care consultations linked to service use and outcome data. *Social Science & Medicine,* 70**,** 473-478.

WILLIAMS, B., IRVINE, L., MCGINNIS, A., MCMURDO, M. & CROMBIE, I. 2007. When "no" might not quite mean "no"; the importance of informed and meaningful non-consent: results from a survey of individuals refusing participation in a health-related research project. *BMC Health Services Research,* 7**,** 59.

WILLISON, D. J., KESHAVJEE, K., NAIR, K., GOLDSMITH, C. & HOLBROOK, A. M. 2003. Patients' consent preferences for research uses of information in electronic medical records: interview and survey data. *Bmj,* 326**,** 373.

WORLD HEALTH ORGANISATION (WHO). 2014. *International Classification of Diseases Website* [Online]. World Health Organisation (WHO). Available: http://www.who.int/classifications/icd/en/ [Accessed 27th October 2014].

WORLDWIDE WEB CONSORTIUM. 2012. *Web Ontology Language (OWL)* [Online]. Available: http://www.w3.org/2004/OWL/ [Accessed 30/06/2014 2014].

WORLDWIDE WEB CONSORTIUM. 2013. *Worldwide Web Consortium Website: Web Ontology Language (OWL)* [Online]. Available: http://www.w3.org/2004/OWL/ [Accessed 27th October 2014].

YE, H. & CHEN, E. S. 2011. Attribute Utility Motivated k-anonymization of datasets to support the heterogeneous needs of biomedical researchers. *AMIA Annu Symp Proc,* 2011**,** 1573-82.

ZHANG, N., CHIN, J., RECTOR, A., GOBLE, C. & LI, Y. 2004. Towards an Authentication Middleware to Support Ubiquitous Web Access. *Proceedings of the 28th Annual International Computer Software and Applications Conference - Workshops and Fast Abstracts - Volume 02.* IEEE Computer Society.

ZHANG, N., YAO, L., NENADIC, A., CHIN, J., GOBLE, C., RECTOR, A., CHADWICK, D., OTENKO, S. & SHI, Q. 2007. Achieving fine-grained access control in virtual organizations. *Concurrency and Computation: Practice and Experience,* 19**,** 1333-1352.

Appendices

Appendix 1. Clinical eScience Framework Policies

Draft List of Practices at CHIME for the Handling of Sensitive Data in the CLEF / CLEF Services Project

1. Introduction:

This document contains a list of common practices undertaken at CHIME when handling the current set of CLEF data, which is released to the consortium under the guidelines of (REFS!!). This document is at an early draft stage and requires input from all relevant personnel.

It should be noted that the policies and practices at CHIME have been loosely guided by the ITU-T Recommendation X.800, Security Architecture Open Systems Interconnection (The OSI Security Architecture) which defines broadly speaking five components for the use of distributed software, essentially Authentication, Access Control, Data Confidentiality, Data Integrity and Non-Repudiation for inter-networking. It is already clear that the requirement for Accountability is not explicitly defined in Recommendation X.800. X.800 has also been used as a recommendation chiefly; The policies do not declare total compliance with X.800.

Consideration of other International Standards such as that of BS 7799 has been taken into account for day to day local data handling - X.800 above does not deal with this domain of communication.

2. Scope:

It is intended that this document forms the basis for the CHIME policy on handling CLEF / CLEF Services data. There are suggestions for implementing updates to the policy written and a number of headings under which the common practices are listed.

3. Data and Associated Assets:

There are five identifiable sets of sensitive information either explicitly anonymised healthcare record data, or by implication, code which allows for access and analysis of the healthcare record data:

- Raw source data from the Royal Marsden Hospital (RMH) which exist as non-parsable XML(REFS) on CD-ROMS
- The same source xml files existing on a computer hard disk or portable hard disk which are both locked away in the secure server room on CHIME and secure office cabinet in the chime research office
- Relationally mapped data existing within a Database Management System (Oracle (REFS) or MySQL(REFS)) on the same computer hard disk or portable hard disk (as above)
- CHIME Source Code which brokers access to the data after it has been relationally mapped
- Record Server and email Public Private Key Pairs Generated at CHIME.

In a collaborative framework, it is necessary to share data across sites and interactions with each

site are documented below. Each interaction comprises two headings:

- Inter-site software communication - the use of software systems developed by the consortium partners to store, broker access, query and analyse electronic healthcare records (these are all described under the relevant inter-site communication heading).
- Ad-hoc data sharing on request - the use of standard transmission means (for example, email, CD-ROM and other mobile media).

Both these communication domains include network communications.

It should also be noted that, at the earliest stages of the project, copies of the CD-ROMS were given to the three other collaborative sites as access to the data was not possible at the time remotely. This included data for diagnoses, narratives and radiology only, and the data was regarded by all parties as being of highest sensitivity, to be neither copied nor distributed.

This policy discussion document is split up into sections based on the locations of all the project partners. They are further organised into sub-sections discussing the two main domains of data sharing, alluding to the specific examples of data assets outlined in (REF). At the end of each section, a set of policies is reduced from this discussion.

4. Flow of Information:

4.1 UCL CHIME:

4.1.1 Common Practice:

Once data is received by Dr Kalra on CD-ROM drives by courier, the data is read either on Dr Kalra's desktop Macintosh computer in his office ('Dipak's Computer') using a Virtual PC version of Windows 2000 (the files are only readable on Windows File Systems), or a secure Windows XP desktop ('Synexius') which is not connected to the network, located in the Research Office room 459 and de-activated after each use.

The data is then sent to Nathan Lea as password encrypted ZIP files using a LAN version of an instant messaging tool, or on USB Keys belonging to Dr Kalra or Nathan Lea. No method involves the data leaving the CHIME network or campus domain. In the event that Mr Lea reads the CD-ROMs first, no local copies are kept on the Windows desktop 'Synexius' and all processing is done in a non-networked environment. The USB Keys are wiped after each use, both directly and indirectly by clearing the trash folder of the desktop where the key is accessed and the contents deleted.

Once the data is opened on 'Synexius', Mr Lea will upload it to the dedicated secure server 'Prometheus' either using the USB Keys only. Once on Prometheus, Mr Lea will run further checks on the source XML to ensure that it is parsable and has not suffered any data corruption through the transfer from RMH onto CD-ROMS, transfer onto 'Synexius' or 'dipak's computer' and then transfer to 'Prometheus'. It will then be imported using the XML import code. The data is processed through the CHIME information markup specifications and record server code, and deposited in a MySQL database, with stringent access permissions. The data is then only accessible remotely provided that the correct code is in the hands of the relevant party, the IP address of the relevant party has been supplied and a roles based access policy is implemented for each remote party at the Record Server(see below). The access code at the server end is co-located with the database which uses a local lookup to access the Database Management System Server.

The data CD-ROMs are kept locked in a filing cabinet in room 459. The original Source files are either kept only on the secure server, or on a portable FireWire Disk, which is locked away in the filing cabinet at all times.

Since September 2004, the database on the secure server has been removed, the source files kept on the FireWire disk, and all source code, Public / Private Keys revoked and the system has only been restarted on two occasions for demonstration purposes. There is no running instance of the access code or RMH data in existence on the CHIME campus.

It should be noted that no source data has been stored on the secure server except when XML processing is required, at which point the secure server is disconnected from the network, the files are loaded from the FireWire Disk, processed and then deleted securely before network re-initialisation.

Suggestions:

- Use only one secure server machine to access data and transfer it to Portable Disk
- Disconnect Server at times of original CD-ROM access and processing for checking XML from any Network
- Import data in a non-networked environment
- Start up network services only when all Server Policies are in place.

Policy Items

- Permanent Records should not, unless under sufficient security lockouts, be kept on Desktop Workstations
- Data transfers should occur either encrypted in a networked environment or unencrypted in a non-networked environment
- Transmission media (outlined below) should not contain copies of data, except for the explicit purpose of transferring data in a non-networked environment
- Should Transmission Media be used for transferring of data, the data should be deleted, securely and immediately
- RMH CD-ROMS, when not being accessed, should be locked away securely at all times, pending total destruction at end of project
- All requests for RMH data from any party should be denied and logged
- Copies of data must be logged
- RMH CD-ROMS should not be exported, copied, transmitted or divulged in any way, either in a non-networked environment or in a networked environment, physically or verbally (with the exception of sharing data, and then under specific circumstances).

4.1.2 Inter-site software communications:

CHIME runs the server software on a single secure server for Record Server use and access to the data, and a query interface. These are subject to stringent access controls implemented using firewall configuration, SSL / RMI and roles-based access and authorisation.

4.1.2.1 The Record Server:

This code allows for the persistence of clinical data in a specific format, marked up with heuristics define by a dictionary or archetype model. This model relies heavily on various persistence methodologies for relational, object oriented and XML database implementations. Components of the Record Server code packages allow for a single instance of the Java Coded Server to access, read, write and update data stored in seven different databases. The versions of these databases have been updated and the record server logic needs to be updated in order to converse with the newer versions of the code databases.

The Record Server is the main service which ties in the code from Sheffield and Brighton, and

client services from UCL CHIME such as the Query Interface (see next section). It has been configured, especially for CLEF, to run as an Remote Method Invocation Service. This means that clients communicate with it using a set of proxies to objectify and interpret the data across networks.

The Record Server is configured to run using Jini, a distributed services registry engine. Jini has been de-activated for CLEF uses of the record server as it cannot provide encrypted communications or distribution of code without major re-configuration. As such, RMI is used for client / server communications over LAN / WAN (see also Inter-Software Communication Process over LAN / WAN), incorporating SSL above the RMI layer to ensure Encryption.

Suggestions:

- Ensure that the Database Management System encrypts data
- Ensure that only authorised processes can access the DBMS.
- Update the Record Server Logic so that more recent versions of the database can be used.

Policy Items:

- All RMH XML parsing and processing should be done in a non-networked environment at the import stage
- All access to, updating or addition of RMH data should be logged by the software processes performing the access, update or addition to the data repository
- All data stored at the database level should be made inaccessible via the DBMS with the exception of the server side software (i.e Record Server and Query Interface)
- The DBMS must possess the facility to encrypt data by the server side process
- All communications with the Record Server must occur using RMI / SSL interface, with explicitly no exceptions (see Communications over LAN / WAN for more details).

4.1.2.2 The Query Interface:

This is a component that allows the CHIME repository to be queried using a more sophisticated set of queries than standard SQL interfaces allow. It can run separately from the Record Server, but should only Communicate over the RMI / SSL link. The only remote machine to have access to the MySQL Database to achieve this was Tony Austin's laptop using the unique DHCP domain name for the machine at both home and office. This was configured through the firewall.

The Query Interface also needed to use the Client Public / Private Keys to communicate with the Remote Server. Clients of the Query Interface, however, do not communicate with it over SSL - the communications are in clear text. This is discussed more fully in the UCL CHIME and Brighton Inter-software Communication section below.

Public / Private Key Sharing and Use:

The fifth set of sensitive data are the Keys that are generated to implement the SSL interface for the Record Server, and the web of trust implemented by Sheffield.

4.1.2.3 CHIME Record Server Keys:

These keys are generated using the java keytool command. The process generates server public and private keys, and client public private keys. The client public private keys and server public keys are passed to the consortium partners in person. The server private key is retained at the server end.

There are mechanisms available which would allow for a more automated approach to distribute the keys, as well as create an manage them. A thorough treatment of key

management is dealt with separately by the CLEF Services proposal. The distribution of secure tokens using standards adhered to by implementations such as Kerberos (REFS) are also a possibility. The current arrangements are however managed too manually and the authentication component relies in this key infrastructure.

Suggestions:

- Make use of Kerberos or alternative session control and token passing implementation
- Make Client keys explicitly identifiable for consortium partners
- Make use of generic software for key management
- Key passwords should not be hard coded into the source code and compiled code distributed to the consortium partners.

Policy Items:

- Under no circumstances shall any CLEF / CLEF Services encryption keys or other PKI tokens be provided to unidentified individuals, groups or organisations, unless they are part of the the CLEF / CLEF Services team and have been authorised for Data Access
- Keys shall have a life span of no more than three months before they are automatically revoked

4.1.2.4 Email Encryption:

For encryption purposes whilst using email as the transmission medium, the consortium has been using GPG - a GNU implementation of PGP keys and alignment with a set of mail clients. The members in the consortium who have generated their own key have digitally signed it and had their key digitally signed by a trusted member of the consortium; this has culminated in a web of trust. There are some consortium members who have yet to produce their own keys for use and this makes it difficult to send sensitive data over email (see sheffield / manchester communications). The responsibility for key management falls to the individuals who have ownership of the keys.

Key holders are:

Andrea Setzer
Catalina Hallet
Dipak Kalra
Adel Taweel
Jeremy Rogers
Henk Harkema
Ian Roberts
Nathan Lea
Richard Power
Stephen Hailes
Tony Austin

Emails which contain excerpts of data or attached data files must all be encrypted. References to data in emails with no direct excerpts from data must also be encrypted in case of accidental disclosure. All emails sent from CHIME which relate to the CLEF data have been logged (See appendix A) and copies are not kept on the email server ('MailSERVER') at CHIME. Mr Lea has kept all of the emails he has composed and received for reference, and will destroy all copies at the end of the project.

Suggestions:

- Ensure all members of CLEF have a PGP GPG key

- Ensure all members know how to use GPG with their chosen mail clients and are confident in its use
- Have a set of keys produced independently of personal key pairs for use in CLEF / CLEF Services
- Have a common CLEF / CLEF Services policy for the management of said keys.

Policy Items:

- All sensitive data shall be sent over email encrypted
- Use of email to send sensitive data should be limited to only absolute requirement
- A register of all emails sent containing sensitive data shall be kept by all parties
- The emails should be kept as part of the logging and project documentation for the project duration and considered sensitive information
- Emails concerning the data must be considered sensitive data.

4.1.3 Working Offsite:

There are rare instances where data work has occurred offsite. One is at the All Hands Meeting in University of Notingham, September 2004, where the consortium handled live demonstrations of the query interface for a limited subset of patients only. The second instance occurs when working from home. Data may be taken if it exists on a laptop computer, or distributed software is tested offsite. Both examples are discussed below.

4.1.3.1 AHM

In early September 2004, at the All Hands Meeting at the East Midlands Conference Centre in Nottinghamshire, a live demonstration was given of Sheffield's and Brighton's software components. No data was actually displayed, and only ITRI's query workbench and Sheffield's information extraction engine was run, with both the Record Server and Query Interface. This demonstration did, however, use the CLEF repository in its entirety at that stage. It was run over a wired, private LAN and full Security lockouts were introduced.

Suggestions:

Use of fake data should be permitted for any demonstrations
No actual data be kept on Laptops when visiting and doing presentations

4.1.3.2 Rare Instances of Working from home:

The practice of taking data offsite, specifically to the home place, is extremely rare. Sensitive data is not stored on laptop machines by practice, and where analysis of data are required overnight or over the weekend, the data is kept on the person and another member of staff will drive them home, or a taxi will be used to get staff member and data to the destination. Mr Lea has worked on the data at home, in a non-networked environment and with the work-room at his home locked, and the hardware locked down. Storage of sensitive data and access to it on home machines is not permitted, and no RMH data is stored or accessed in such a fashion. The use at home of Laptops from work containing sensitive data is the scenario in rare cases, but strict policies govern the home networking environment and practices at the home place were in effect, including encryption of the laptop hard disk (using a tool called fire-vaulting on the Mac OS X). For the majority of the project duration, there has been no sensitive data of any kind taken to the home place and then only by Mr Lea. The data is not stored there, either on disk or other removable media.

Emails concerning the project have been answered from home, using the security methods outlined above.

Suggestions:

- Limit instances of storing sensitive data on Laptops
- Never store any sensitive data at home
- Have all access logs at home available for scrutiny

Policy Items:

- Data should in no way, manner or medium, be stored at home or on home systems
- Offsite backups may be stored at home so long as they are encrypted, password protected and locked away in an appropriate, locked cabinet or safe
- Laptops must not hold any sensitive data (see above)
- Remote access to Secure Servers based at CHIME must not occur, unless within the security frameworks specified, but never from the home place
- Work on sensitive data at home or offsite must not occur outside the work environment
- Any demonstrations must use false data
- Presentations and slide shows must not contain any sensitive data whatsoever
- No sensitive data will at any time be transfered to another person's laptop or media for any reason whatsoever

4.2 UCL CHIME and Sheffield:

4.2.1 Inter-site software communications:

Sheffield has produced one Software component which has interacted with the CHIME services over the RMI / SSL interface. This is an information extraction engine which downloads RMH data to the local Sheffield server. The code analyses the downloaded data, extracts relevant pieces of information, and then sends the extracted elements back to the CHIME repository over the SSL / RMI link. Local copies of the data used during the analysis are deleted during this process.

The software communication across this site must occur using the Inter-software communication process over LAN / WAN documented in section ????

Suggestions:

- Have repository side code run the extraction and information upload so no data is actually sent to a remote site.
- Have all communications occur over a secure VPN (for example IPSec)

Policy Items (see also Inter-Software Communication Process over LAN / WAN):

- The LAN / WAN communications must occur with standardised encryption methods (including SSL or TLS, for initial phases, with a move to IPSec)
- The communications must rely on an authentication mechanism (such as proposed by the PKI standard, for example.)

4.2.2 Ad-Hoc Communications:

It has been necessary to send Sheffield updated code fragments as well as raw data. These also include new public / private keypairs for the CLEF client in Sheffield.

In the event of any such software or raw data being sent, attached files have been encrypted using the recipients' public key, and the body email has also been encrypted through the standard mail clients (Apple Mail, Eudora, Outlook amongst others). It has been necessary to encrypt body mail and attachments separately as certain mail clients use ????? which will not

necessarily send attachments encrypted.

Use of FTP and other such protocols (including HTTP) have been used, but only for encrypted data; use of both HTTPS and SFTP on several occasions.

Raw Data, relationally mapped data and code elements have never been communicated to the Sheffield group using CD or DVD-ROMs. This practice should continue.

Raw data, relationally mapped data and software components have, on very rare occasions, been shared using USB keys. This is only when the exchange partners have been in each others presence, and it is ensured that the data is removed from both USB Key and trash folders on the target machine. The data is encrypted too.

Raw data, relationally mapped data and any code access is explicitly forbidden from being put on the Twiki in any communication. UCL CHIME have never done this.

The security elements of the Twiki need to be examined and implemented to a higher state of security.

Suggestions:

- Curtail use of USB Keys for data transfers, even if done locally and kept on keys for a very limited period of time.
- Curtail transmission of any unencrypted data over any means using any protocol.
- Define a list of Mail Clients which adhere to appropriate standards

Policy Items (see also Ad Hoc Communications):

- No data shall be exchanged on any portable media including CD-Roms, DVD-Roms, Flash memory components of any kind (including flash memory cards, USB Keys etc) under any circumstances (with the possible exception of encrypted data, which must be logged).
- HTTP and FTP should only be used for transmission of encrypted data.
- Access logs for HTTP and FTP servers must be kept for the project duration.

4.3 UCL CHIME and Brighton ITRI:

4.3.1 Inter-site software communications:

Brighton has produced a set of code which provides a Java Swing GUI to construct queries for researchers to pass to a generic Query Interface at the UCL CHIME Repository end. This communication submits queries and retrieves figures, sending them back to ITRI's tool. This happens over RMI - there is no SSL interface as there is no actual patient data being sent across the communications link. There are of course search criteria transmitted to the Query Interface and numerical results returned.

Suggestions:

Implement the SSL interface as soon as possible.

4.3.2 Ad-Hoc Communications:

There have been limited instances of providing code to Brighton. No raw data has been exchanged, with the exception of the start of project (see Section FLOW OF INFORMATION).

Code is sent to Brighton using Encrypted Files sent in encrypted emails (as with Sheffield). On one occasion, Dr Kalra Distributed an encrypted set of code libraries to Brighton.

4.4 UCL CHIME and Manchester:

4.4.1 Inter-site software communications:

There has been no Record Server code exchange or inter-software communication between UCL CHIME and Manchester. There have been requests for code to be uploaded to a centralised server, which have been denied by UCL CHIME.

On one occasion, server side PHP scripts have been used by Manchester, and Manchester only. The PHP scripts ran on the Apache Web Server and were accessed by Dr Jeremy Rogers. The server had a firewall configured to allow only the machine used by Dr Rogers to access it remotely, and only on port 443 (the default for Apache's SSL interface). The server also allowed only Dr Rogers' IP address (which is static) to access it, and Dr Rogers had to authenticate using Apache's log in screen. While this usually sends data in plaintext, the entire transaction was handled using SSL via Apache). SSL was used for the whole interaction which occurred on one afternoon. The data accessed was limited to only a select few queries from a co-located MySQL Database. This service was shut down at the close of business on the day of use (15/12/2004).

4.4.2 Ad-Hoc Communications:

Chris Garwood at manchester was asked to inspect and parse the RMH CD-ROMs specified in section ??? FLOW OF INFORMATION, specifically the Narratives and Histopathology, to tidy up the unstructured text format in the narrative components. Data was returned on encrypted on CD-ROMs or encrypted email.

Suggestions:

- Continue Policy of not allowing Manchester to keep a repository of code
- Do not re-activate PHP interface unless in extreme circumstances
- Curtail sharing of CD-ROMs between partners as this is no longer necessary.

5 Inter-Software Communication Process over LAN / WAN:

Communications between sites running the collaborative software require the use of either a Local Area Network or a WIde Area Network to send and receive data packets. A process (REFS STANDARDS) occurs whereby data is encoded in a form which can be transmitted using physical media (such as fibre optic cable). The data is then decoded so that this is readable at the transmission endpoint.

Without the use of encryption, it is possible to decode the data stream between the two communicating sites. There are commercial and open source tools that are built to monitor network traffic (REFS). This makes LAN and WAN communications insecure and inappropriate for the transmission when data is unencrypted. It is for this reason that SSL was chosen to implement encryption for network transmissions of data during the software communications process. Data which is transmitted and routed is encrypted, so that in the event of unknown monitoring (either routine or unauthorised) over networks where the participating consortium partners have no real jurisdiction and do not ensure Quality of service, can only be read at either end of the data transfer.

The SSL implementation at the Record Server is not, however, the standard SSL implementation.

There is an amendment in its implementation for the Inter-software communication process over LAN / WAN which requires that a set of public / private key pairs be distributed to the partners before any data transfer can take place. This is outside of the standard SSL implementation which uses a symmetric encryption methodology at transfer time. The CHIME SSL implementation adds a degree of authentication to the SSL implementation which standard SSL is not equipped with. It remains uncertain though as to how secure either the key distribution or key storage post distribution actually is (SEE ABOVE).

The implementation of a private network over WAN communications is also under consideration. This involves the establishment of Virtual Private Network (VPN). There are authentication and encryption implementations for Secure VPNs, utilising the IP Security standard, or IPSec (REFS). It would effectively be a closed network accessed by the consortium partners who had the the software installed and the authentication details required.

The network communications are but one layer for consideration. The access to and maintenance of the software accessing the network and transport layers are of great importance. Various implementations at different levels are available to authenticate and provide access to different software components. One such example is that of session management for entities (human or software related). Mechanisms exist in Web Service enabled software, for example, to allow access to assets for limited periods of time based on an authentication and thereafter role based access provisions(REFS)

Suggestions:

- Implement a tighter authentication component for CLEF Services inter-software communications over WAN / LAN, making use of appropriate authentication and session management implementations
- Implement standardised token exchange and management using Kerberos or other software.
- Consider the possibilities of removing SSL implementation and using IPSec for encryption and authentication delivery.
- Consider the possibility of using both IPSec and SSL
- Consider the issues of Kerberos implementation
- It must be possible to audit and trace any network communication attempt whether successful or not.

Policy Items

- The LAN / WAN communications must occur with standardised encryption methods (including SSL or TLS where appropriate, but based on PKI and AES)
- The communications must rely on an authentication mechanism (using PKI, AES as well as network layer authentication) at the endpoints for communication
- Communications should occur over a private network (a VPN) with the appropriate Security standard (IPSec, for example)
- All network traffic must be logged at the endpoint server end using default logging tools. Logs must be retained for at least the project duration, and up to seven years after that (REFS).
- All network communication attempts must be logged at the Client and Server endpoints, be they successful or otherwise
- All network communications which read, write or amend data must be logged at the client and server endpoint (see Record Server and Query Interface discussions)
- All network communications which gain access of any sort to the server machine, server software or data repository must be logged at the server endpoint (see Record Server and Query Interface discussions).

Appendices

Assets Register

The following machines are used for accessin and managing CLEF and Project Related Data:

Prometheus: Apple PowerMac Dual G5; Apple Macintosh OS X 10.4.2 Server Edition: Main Record Server and Query Interface Server Machine. Contains relationally mapped data, server keys and experiment result data

Alchemy: Apple Mac Power Mac G4; Apple Macintosh OS X 10.4.2: Desktop used for some experimentation and storing small amounts of data for limited periods. All data removed from this machine in October 2004 and there has been none since

Pangu: Apple Powerbook Laptop G4; Apple Macintosh OS X 10.4.2: Laptop used as Mr Lea's primary computer tool. Some Data stored for AHM - relationally mapped only. Other times include source files which are securely deleted after used. No data resides on this machine in any form.

Synexius: Dell Dimension with Windows XP: Primary research machine for initial code tests. Some data exists but Synexius is inaccessible over the networks at chime and it is locked away in Room 459. Pending removal of data and Hard Disk destruction.

Nuclear: Dell Server with Windows 2000: Used to hold relationally mapped data for 50 patients in Oracle 8i Database. All data and experimental data removed and securely destroyed.

Quandary: Dell Optiplex with SuSE Linux: hard disk failure in certain sectors. Disk to be destroyed and replaced. Machine in room 459 securely.

dipak'scomputer: Apple IMac running Mac OS X 10.4.2. Contains Source Files. Stored in Dipak Kalra's office.

Lacie Big Disk 250GB: Contains Source and Relationally Mapped Backup Files. Also all code and Key Backups. Locked in Filing Cabinet in 459. Mr Lea has the keys.

Original Source CD-ROMs are all locked in the cabinet in 459.

1 Backup DVD-ROM is also there in the filing Cabinet containing relationally mapped and compressed source files.

Repositories Installed at Consortium Sites:

Statistical Disclosure Control, Cathy Marsh Centre, University of Manchester:

1 Mac Mini with complete set of Database, Record Server, Query Interface code; RMH Source Files (including pre and post regular expressions versions).

Security Arrangements: No Network Connectivity, mac mini stored in locked up server room used in the past for analysis of Census Data

2 University of Sheffield: Record Server, Query Interface and Database records. NO RMH Source Files.

Security Arrangements: Stored on an encrypted drive on Ian Robert's Laptop. All data is mounted when used, but invisible and encrypted when not in use. This is sufficient for now

and allows use on a Laptop. All other restrictions apply according to Phase 2 policy.

3 Open University: Record Server, Query Interface and Database records. NO RMH Source Files.

Security Arrangements: Stored on a dedicated secure server under the management of the IT group in Milton Keynes. All team members have signed CLEF confidentiality agreements. Directions according to phase 2 requirements have been followed. IT Staff are competent and knowledgeable. Remote access is done by SSH only. Maintenance of Server should be adequate given staff expertise.

Security Policy Document - Phase 2 Deployment

Introduction:

This document defines a basic Information Security policy to allow CLEF consortium partners to access the repository of complete Royal Marsden Data Sets, completed 22nd May 2005.

Review:

The next review of this policy should be scheduled for 3 months time. Further consultation with UCL IS, CERT and CHIME System Administrator should occur at that point.

Scope:

This policy defines a network access for Brighton/OU and Sheffield NLP. Manchester SDC is covered for local, non-networked access. This policy document also describes the server side repository policies.

UCL-CHIME - Repository Service:

The CHIME repository is composed of a MySQL database management system which stores the relationally mapped data and the Java Record Server which allows access to the database. Intrusion detection systems (SNORT) are deployed, and standard system level logging occurs. There are two repository level logs: the RMI Service log and database log.

All access to the repository must occur using the RMI/SSL interface.
No direct database access shall occur, except via the record server.
No other services, except for SNORT, will run on the Server Machine.
A firewall must be in place - contact to only the RMI registry port and Record Server Port is allowed, only from specific machines, identified by IP address.
The machine must be patched at least weekly or in the event of extraordinary requirement.
Logs must be checked daily.
Encryption keys must be updated monthly.
Antivirus software must be installed and run every two days.
The service shall remain online on weekdays between 8am and 6pm.
Wireless networking will not be used.
Dial-up access to networks will not be used.

Sheffield NLP:

Sheffield will download narrative data, scan it and add information extraction and chronicle data back to the repository.

Sheffield will access the Record Server only on a newly installed operating system with firewall controls allowing access for RMI transactions and the Record Server only.
Access to the record server will occur only over the encrypted SSL interface.
No other services will be run on the accessing machine.
The machine must be patched at least weekly or in the event of extraordinary requirement.
Antivirus software must be installed, updated and run every two days.
An Intrusion Detection System with logging facilities must be installed and running.
Audit logs both at service and system level must be made available for scrutiny where required.
No copies of data shall be made or distributed on any medium.

CHIME code will not be copied or shared in any way, either with individuals or other machines.
Only the one machine shall have access to the services.
No access shall be allowed from home or other sites.
Wireless networking will not be used.
Dial-up access to networks will not be used.

Manchester:

SDC:

The Cathy Marsh Centre shall require access to the repository on a non-networked basis only. To that end, a Mac Mini shall be provided by CHIME, with the repository and source files placed on it.

The Mini shall be locked away in their secure centre.
Data shall not be placed on any other machine, copied or transmitted from the Mac Mini.
All CMC code shall be run on the Mac Mini and results stored therein.
The Mac Mini will not be networked.
Wireless networking will not be used.
Ethernet networking will not be used.
Dial-up access to networks will not be used.
Firewire Networking will not be used.
Bluetooth will not be used.
Backups may be made by SDC but they must be held in a locked, non-accessible format, especially on another machine. A DVD-R drive is provided on the Mac Mini so that DVD backups may be made. These are to be destroyed when the interaction with the Mac Mini has finished.
Sophos Antivirus will be run every week and the logs will be kept.
Networking for the sake of patches and Sophos updates is also not permitted, except in an extraordinary emergency, at which point a firewall policy must first be agreed.
Audit logs from system and software will be retained and in no way tampered with.

Jeremy:

Jeremy will need access to a shared repository to work with Sheffield. The requirements are similar.

Manchester will access the Record Server only on a newly installed operating system with firewall controls allowing access only for RMI transactions and the Record Server only.
Access to the record server will occur only over the encrypted SSL interface.
No other services will be run on the accessing machine.
An Intrusion Detection System with logging facilities must be installed and running.
The machine must be patched at least weekly or in the event of extraordinary requirement.
Antivirus software must be installed, updated and run every two days.
Audit logs both at service and system level must be made available for scrutiny where required.
No copies of data shall be made or distributed on any medium.
CHIME code will not be copied or shared in any way, either with individuals or other machines.
Only the one machine shall have access to the services.
No access shall be allowed from home or other sites.
Wireless networking will not be used.
Dial-up access to networks will not be used.

Brighton / OU:

Brighton/OU will be testing two query workbench tools. One is the standalone Java application. The other is the same application programmed in applets. The applet version will have to access a dud repository. There will be access from Catalina at home over a broadband connection. This in turn takes the use out of the JANET network perview.

Brighton / OU will access the Record Server only on a newly installed operating system with firewall controls allowing access only for RMI transactions and the Record Server only.
Access to the record server will occur only over the encrypted SSL interface.
No other services will be run on the accessing machine.
An Intrusion Detection System with logging facilities must be installed and running.
The machine must be patched at least weekly or in the event of extraordinary requirement.
Antivirus software must be installed, updated and run every two days.
Audit logs both at service and system level must be made available for scrutiny where required.
No copies of data shall be made or distributed on any medium.
CHIME code will not be copied or shared in any way, either with individuals or other machines.
Only the one machine shall have access to the services.
No access shall be allowed from other sites.
The Applet code shall not access the repository. Only a dud repository need be used.
Wireless networking will not be used.
Dial-up access to networks will not be used.

Audit:

Audits will be scheduled regularly. It is proposed that a committee be set up from within the Clef consortium to facilitate this.

- All sites must make themselves available for audit.
- Conformance to the above will be checked along with audit logs at each site.
- System and software logs must be made available.
- Audit will be undertaken by the assigned committee and / or another consortium partner.

POLICY REVIEW:

It is proposed that this policy be reviewed in three months' time. A formal review group will be established within the next three months.

Appendix 2. CLEF Roles and Privilege Management Results

Roles Specification and Security Authorisation Description							
WYSIWYM Query Workbench		A tool that allows for query construction in natural language and submission to the repository. Aggregate data is always shown and provided to Clinical Researchers					
Sub-Role		**Functional**	**Non-Functional**	**Risks**	**PERMIS Access**		**Component**
Local Access Broker		Request from Repository Administrator to gain access to the Repository	Conversation: email; telephone	Disclosure due to conversational requirements. Communications taps			
		Receive proxy stibs for RMI access	Email encrypted code (receipt)	Theft of code could allow a masquerader full access to Query Interface if they understand the manner in which they must sibmit queries (Javadocs online).			
		Ensure that connection works in liaison with Repository Administrator	Repository Administrator sends code stubs and Catalina integrates	Possible masquerading of Catalina to gain access to repository.			
		Liaise with Repository Administrator to allow for functionality enhancements		Updates to code may have a knock on effect either way according to need. Older versions of the code may still present an unauthorised access risk. Back end may have effect on front end interaction requiring code update.			
WYSIWYM Developer		Run Queries using WYSIWYM editor	Submit Queries to Query Interface	Possible disclosure risk based on analysis of queries and agregate results			
		Interpret and display results	View data on WYSIWYM; if exception is thrown, liaise with Repoaitory Administrator	Data must be accurate wrt ensuring that the software presents data accurately			
		Run test scripts for testing result sets					
		Proxy to end user in terms of building as per reasons to begin process in the first place.					

Table 31: Roles specified for Clinical Query Workbench Handler

Roles Specification and Security Authorisation Description						
Repository Administrator		General responsibility to ensure that Repository and Query Interface Services are running effectively and appropriately; Responsible for Doing EHR Record Imports				
Sub-Role		**Functional**	**Non-Functional**	**Risks**	**PERMIS Access**	**Component**
Data Accuracy Checker		Ensure repository provides accurate and correct data	Use mysqlcheck to ensure integrity of database	Full access to data necessary; high trust level required	READ	Relational Database
			Run maintenance checks of server software and hardware.			
		Ensure that data imports do not scew the integrity of the data in terms of quality of source data, accuracy of import and effectieness of database storage.	Run checks directly on the database using SQL	Full access to data necessary; high trust level required; access to all Record Server and Import Code required, as well as database	READ	Relational Database; EHR Record Impoerter; Record Server
			match results to Configuration Specification of bindings; Run mysqlcheck after imports			
User Request Manager		Handle requests from Accredited Researcher, IE Importer or relevant technical staff to access data	Provide Accredited Researcher with access keys for repository;	Access to repository is in the hands of this one role by means of certificate exchange: may not require access to data, but by default, will have it via Services	??Is this a certificate manager task now?	Record Server Code
			Liase with Accredited Researcher or Technical Assistant to ensure that repository is accessible as a service with correct coding scheme	Prone to emergent behaviour between parties: corner cutting, horse-trading for the sake of practicality or something more sinister	READ	Record Server
		Handle any queries that may arise as a result of Acredited Researcher interaction with data				
Auditor (general or Repository Administration)		Ensure that information security policy is adhered to as per specification from Security Administrator	Check information security policy	Prone to Draconian attitude;	READ	Service and Server Logs
			Prove that implementations match policy requirements			
		Ensure that Roles are adhered to as per specification issued by Roles Administrator		PERMIS' job!		
Service Provider		Use sensible metrics to guage when repository should be accessible	Take repository offline and power down server when it is clear that access is not required	Data not available at a reasonable time to ensure that Accredited Researchers can work; prone to over zealousness		
			Leave server off network when running overnight checks and administration	False sense of security		
		Ensure that when live, repository behaves as expected and meets needs of researchers	Monitor communications, access and activity logs and use SQL to provide a frame of reference; use database access directly to deal with any issues that may arise			
		Arrange and administrate Backup and disaster recovery	Clone database, code and access files for services	Another resourcce to protect	Read, pull, write	Record Server
Developer		Update service source code as and when required within bounds provided by and information security policy, agreed requirements for services and running requirements of associated services	Use appropriate tools for software development; ensure rollback is possible to provide a working and reasonably assured secure service			
		Ensure within reason that updated code is tested, validated and debuged; all security risks must be met as far as is reasonable	Adopt appropriate testing methodology for code; establish appropriate metrics for security tests; use sensible approach for debugging and code generation			
Record Importer		Oversee or enact data imports to repository from source data provided by Clinical Data Provider and anonymised by CLEF Identity Gatekeeper	Use EHRImport service to bind source data to Archetype Model	Total and unfettered access to source materials and stored data; import process is slow and	WRITE	EHR Record Importer; Record Server

Table 32: CLEF Data Repository Administrator Roles

Sub-Role		Functional	Non-Functional	Risks	PERMIS Access	Component
SDC Analyst		Release a set of cross classiication of tables of data to see if they can find out how to infer details of people in the repository				
		Look at cross classification from set of information, and condition down to a sub population based on knowledge about some person in the repository: this is true if they all share something in common		Generates a lot of cross tabulations of the repository - need to look at any part of the data so that there is a maximum level of trust; May need to be involved in actually identifying the patients; doesn't matter what indiation is conditioned on - there is limited certainty		
		Take into account prior queries				
		Exact inferences are the issue				
		Infer 0 in cross tabulations to see if identification is possible				
		take cross tabulations of multiple variables				
		keep a track of information that is released				
		Look at what a user has requested and are requesting to see if current release is safe				
		search for maximal set that can be released				
		Check queries based on variables for cross tabulation, and any other queries related to that would be ok.				
		Analyse data				
		Cache previous requests				
		Maintain and query local set		Synchronisation update would be needed as well from main repository. This would be onerous to ensure consistency, as well as the possibility of accidental leakage in the process of update		
		Screen for possibility that the may get back to the counts.		Inferences are likely to be wrong as well in a perfect world. Narrow down to one patient. Possibility of exact match should be considered.		
		Attack scenarios need to inform on where the attacks and sensitive items are likely to be				
		Assume that there is no collaboration		Not necessarily realistic?		
		Do not add noise to data and data quality - IE has added noise.				
		Software that allows entering of arbitrary tabular data and get some bounds back				
		Deal with updated data sets because new individuals added		When there are many, it is ok, but only one would be problematic and risk identification of that one addition		
		Check for query similarities				
		Users to choose aggregation of variables (WYSIWYM Editor users)				
		Limit users appropriately from specific queries after use		THIS MAY BE TOO RESTRICTIVE		
		Sensible attack scenarios				
		Try and get an expert opinion on how to attack this data and generate fake data				
		Use test script to test methods				
		Do one table at a time to deal with this analysis				
		Sensitivity of variables				

Table 33: CLEF Statistical Disclosure Control Analyst Roles

Roles Specification and Security Authorisation Description						
Chronicle		An idealised model of a patient - a single entity representation with cross references. This component involves a Chronicle Analyst, Chronicleiser, an Ontologies Maintainer as well as technical and development teams				
Sub-Role		**Functional**	**Non-Functional**	**RISKS**	**PERMIS Access**	**Component**
Chronicleisation		Chrionicleisation process will link in Information Extraction process and results		The inherent risks are much the same as Information Extraction		
		Process will access the repository directly		This will mean that the process will need unfettered access to all data, except perhaps the narratives...	Read	
		Process wil push and pull data - will need to read and write		This has ramifications in terms of data integrity, as well as local arrangements for downloading and analysing data	Read, Write, Push, Pull	Chronicle Service; Record Server
		Maintain locally stored data		This makes the Chronicleiser just as sensitive as the repository.	Pull	
		The data is represented as agregate data and will be stored as actual records, as a summary		Makes the Chronicle service as sensitive as the repository		Chronicle
Chronicle Analyst		Use inferencing to derive chronicle contribution to main repository		Inferring new information from the data		Chronicle
Chronicle Manager / Administrator (Could be Analyst)		Will need to identify patients to check that chronicle is working as well.		Full identification required	Read	Chronicle; Record Server; Relational Database?
Clinician Administrator		Access Chronicle		Infer components from handling the data		Chronicle
Ontologies Maintainer		Define and maintain ontologies for use to identify clinical sepcifications		Geared towards chronicle which is geared towards the repository, so the ontology maintaioner would have to eyeball the data		Chronicle; Record Server; Relational Database?
Chronicle Developer		Develop Object Model for Chronicle process		Another Developer who needs to have access to the original data		Chronicle; Record Server and Relational Database?

Table 34: CLEF Chronicle Operator Roles

Roles Specification and Security Authorisation Description								
Auditor								
		Responsible for the auditing of all access and interaction with CLEF repository. This spans external and internal interactions, be they within the CLEF Consortium or from external users wanting to access CLEF resources						
Sub-Role		**Functional**	**Non-Functional**	**RISK**			**Permis Access**	**Component**
Audit Clerk		Run / view Exception Reports	Check logs of services for access					
		Run / view Activity Reports	Check QI and RS logs to monitor activity. Check counterpart logs of clients					
		Run / view SDC / Query Vetting and Analysis Reports						
		Investigate potential disclosure						
		Detailed Patient Record Investigation	Check the kinds of access that have occurred in direct data access either to database or RMH Source Files					
Auditor		Allow access to Audit Clerk						
		Investigate non-routine investigations of potential disclosure and detaile patient record investigation						
Audit Manager		Authorise investigation (and assign / revoke rights						
		Adjust trigger settings for Exception and other Audit reports	Not yet built the functionality to allow this.					
		Audit' audit activity						
		Run/View (medium term) activity reports (including those of Audit Clerk and Auditor)						

Table 35: CEF Auditor Roles

Appendix 3. DHICE Common Policy Framework

Generic Information Security Policy Template

Version History (a change history of the policy document – this should be completed as the document evolves):

DATE	VERSION	AUTHOR	COMMENTS

Corresponding Author:
Mr Nathan Lea
Centre for Health Informatics and Multiprofessional Education
4th Floor, Holborn Union Building
Highgate Hill
London N19 5LW
Tel: 020 7288 3798
email: n.lea@ucl.ac.uk

1. Introduction and Background Details

Preamble

This document represents a suggested generic template for information security policies that are produced to inform working practice and document stipulations for the use of medical data for purposes other than those for which they were collected. These include:

- ethically approved, research council funded research projects performed by research institutions;
- Public Service departments that perform disease surveillance, statistical processing and commissioning activities.

The generic template has been compiled with reference to and consideration of a series of documentation that it is worth reviewing that includes:

- ISO 27000 Series (British Standards Institute, 2005a), most pertinently ISO 27001 (British Standards Institute, 2005b) 27002 (British Standards Institute, 2005a) and 27799 (The International Organisation for Standardization, 2008) – the series is available via UCL Library Services (and most likely your institution's agreements with ISO);
- Connecting for Health – Information Governance (Connecting for Health, 2011) – this is a very useful portal that offers access to NHS standard documents and guidelines for a wide range of uses, from clinical care, data curation and for research;
- The Information Commissioner's Office (Information Commissioner's Office, 2014a) offers a wide range of details, including data protection (Information Commissioner's Office, 2014c), technical guidance on determining what personal data is (Information Commissioner's Office, 2011b), what data derived from personal information is pertinent for the Data Protection Act (Information Commissioner's Office, 2011c) and a code of practice for sharing data (Information Commissioner's Office, 2011a);

- Institutional guidelines and policy frameworks exist as well. UCL is in the process of consolidating it's online guideline and governance requirements documentation (University College London, 2010a) by way of example (see your institution's documentation repository).

It is worth pointing out that the guidelines and good practice recommendations have recently specified the need for more documentary resources, which include more detailed specification of practices and sharing by organisations and initiatives handling sensitive data resources. The recommendations suggest documentation that includes details that should already be included in the information security policy but as part of other agreements and contractual obligations. Thus the information security policy, it would be fair to say, is very much more clearly comprised of a set of documents or a framework of resources rather than one single document, the policy being the overarching resource that should contain reference for the related documentation.

The policy itself:

The introductory section of your policy should include:

- concise details about the project, how it is funded and an introduction about what the project is trying to achieve;
- if the project uses a particular programs or databases that has been built specifically for the project, an introduction on this / these and an overview of how they work;
- if the project collaborates with other partners or projects, a concise description of these (to be listed in a separate section);
- details about the legal framework within which the project operates and permissions that have been granted from ethics committees or other statutory bodies;

- details about how the policy document has been established, who has been involved in its creation, the guidelines and standards used for its establishments and who / what has contributed;
- details about whether this policy meets, replaces or should run in parallel with any other policies or guidelines that are in existence;
- details about reviews and timescales for reviews of the policy.

2. Organisations, Members and Resources Involved with the Project

This section should list named members of the project, the organisations involved with the project and their roles within the project. As with the details outlined in Section 1, these could include lists, or reference to where the lists are stored.

There should be details here about those involved with the creation and stewardship of the policy.

There should be details about the other policies that might be in operation throughout organisations and projects that are involved with the project, particularly if different policies relate to specific data items and resources covered by this policy document.

There should also be a detailed list of the assets that are being used for the project, including:

- Data items;
- Hardware resources;
- Software resources;
- Transmission and dissemination resources.

This section should also include details about the activities that are to be performed with the resources, and then cross-referenced with a list of users who would be undertaking the activities themselves, which should also appear here.

There should be a list or reference to a list of any third party suppliers who have a role in maintaining the resources; they should be already be identified in the list of users.

Where users are members of staff, there should a reference to where their contract of employment is stored, with whom they are contracted and for how long. There should also be an indication as to whether the contract specifies penalties for violating the contract where the processing of data protection is and confidentiality is concerned, or the policy itself should specify penalties. You may find it useful to have a list of previous offenders and penalties that were bestowed upon them.

Where agreements are reached between organisations and projects to share data resources, this section should also include reference to data sharing agreements between collaborating partners. The agreements themselves could either form part of the appendices but should at least be referenced by the policy document(s) and form part of the policy documents' framework. See the Information Commissioner's Office recent data sharing guideline document (Information Commissioner's Office, 2011a) for precise details of sharing stipulations and agreements between parties.

Increasingly since 2010 your project might have had to enter into data sharing agreements with data holders so you should to refer to the stipulations therein (or again include the documents in the appendices / part of the document framework).

Intellectual Property Rights are another element that has garnered increased attention over the last decade. It may be worth including reference to documentation that defines these provisions in the policy, though it is not currently explicitly required. Licensing issues are becoming more topical as the

expectation of funding bodies is increasingly about making available information and resources.[1]

3 Risk Assessment and Analysis

Good practice guidelines specify that a risk analysis should be run. Details of this analysis should be described here, and based on the details in Section 2. The analysis usually includes the risks associated with the use of each asset in the context of the activities that are being performed with the resources: the focus of the analysis is to determine the asset vulnerabilities, the likelihood that a threat will take advantage of a vulnerability and the cost should the asset be compromised. This will inform the following section, where activities are listed again and stipulations are specified as a series of controls to mitigate risks as the activities are carried out.

There are a number of risk analysis approaches, including SABSA (Sherwood et al., 2005) and the Information Risk Management pages on the CfH website for some examples. The activities and outcomes of the analyses and assessments should be detailed here.

4. Activities and Stipulations for Securing Resource Use

This section should list the activities identified in Section 2 along with the resources that are being used to carry out those activities and the stipulations that specify the control measures that should be asserted. For Example:

"Data Transfer between collaborating site A and B: this involves the sharing of dates of birth for data subjects that are part of the project cohort between two sites so that they can run analyses on these dates of birth. The transfer is bi-directional, and should be achieved using the secure "Document Transfer" system that has been set up by A and B's organisations. Dates of birth are exported to a

[1] Whilst the licensing particulars are arguably beyond the scope of this document, the activities for

CSV file, and prior to transfer via the "Document Transfer" the files should be encrypted using a 256bit AES encryption cipher. The key should be symmetrically encrypted, and the password should be given to the recipient by telephone after they have been identified."

Where systems that use security components like access control are part of the control assertion, the configuration of that system should be specified either in this list of activities, or a reference to the configuration specification should be made if it is not included here.

5. Appendices:

These could include copies of or reference to data sharing agreements.

There should be a list of abbreviated terms and acronyms, particularly if they include project specific terms. For example: "AES – Advanced Encryption Standard"

A dictionary of common terms might also be appropriate for clarification.
"Advanced Encryption Standard (AES): Encryption Standard that uses block ciphers to… ref: http://en.wikipedia.org/wiki/Advanced_Encryption_Standard "

Appendix 4. UK Community Advisory Board Presentation

Databases for HIV: Integration Collaboration Engagement

UCL

Patient Information Security

An overview of practice and procedure

UK CAB Meeting
13th April 2012

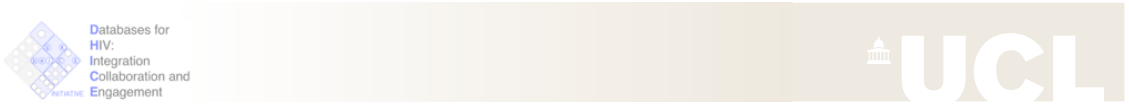
Nathan Lea
Senior Research Associate
CHIME, UCL

Databases for HIV: Integration Collaboration and Engagement

UCL

Overview - Questions that have been asked

- What happens to information collected about me, who has access to it and how is it used?
- How is it protected?
- Further information



Background

- What happens to information collected about me, who has access to it and how is it used?
- Information is collected about you to provide care services and support care decisions
  - includes demographic information, lab results, co-morbidities
- De-identified clinical information valuable for research
- Also informs:
  - population health surveillance (including condition prevalence)
  - healthcare policy and strategy
  - commissioning of services

How is information protected?

- Core principle across both clinical care and secondary use environments
- Can't go into explicit detail! BUT this involves:
  - Development of Information Security Policy
  - Risk Assessment and Analyses
  - Applying protection mechanisms in practice

Databases for HIV: Integration Collaboration and Engagement

UCL

Information Security Policy

- Defines management and user responsibilities
- Guidelines on how to handle information securely (based upon mitigation strategies)
- Most highly regarded international standards - the ISO 27000 Series
- Several guideline documents that offer additional guidance within the NHS
- Information Commissioner's Office - data sharing agreements

Databases for HIV: Integration Collaboration and Engagement

UCL

What do policies contain?

1. Introduction and Background Details of Organisation handling information

2. Organisations, Members, Service providers and Resources Involved with the Project (plus details of any data sharing agreements)

3 Risk Assessment and Analysis

4. Activities and Stipulations for Securing Asset Use

Databases for HIV: Integration Collaboration and Engagement

UCL

3 Risk Assessment and Analysis

- Identification of information assets (records, databases, servers, disks etc.)

- Vulnerabilities - weaknesses of the assets exposed when used (portability, accessibility, value…)

- Threats - aspects that can exploit a vulnerability to attack an asset

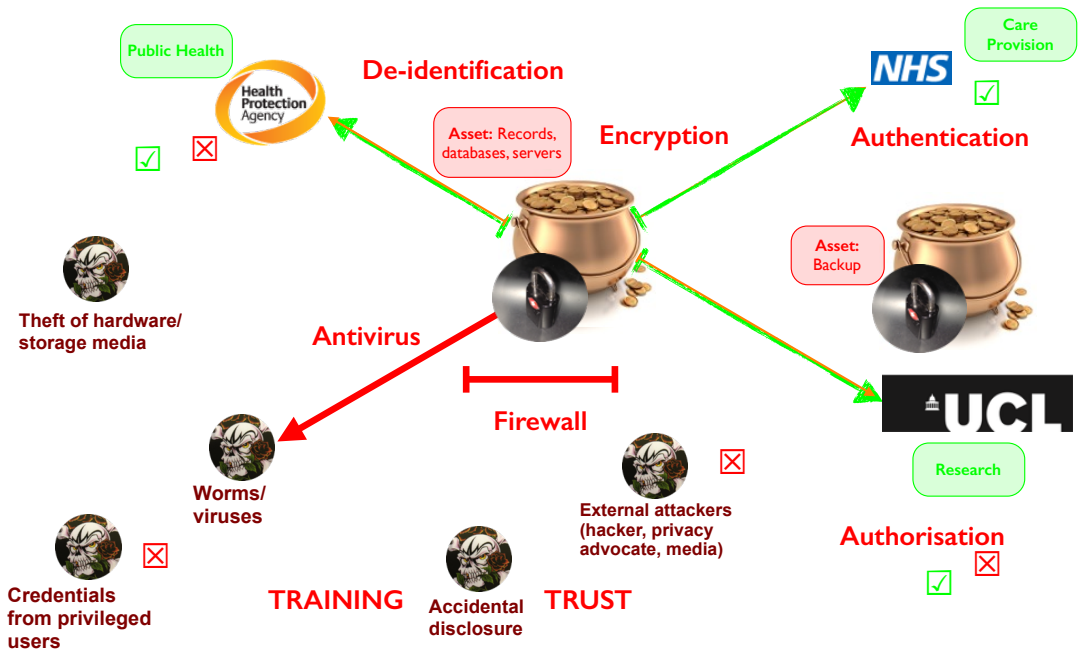
- Risk assessment - the likelihood that a threat exploits a vulnerability against the potential impact…

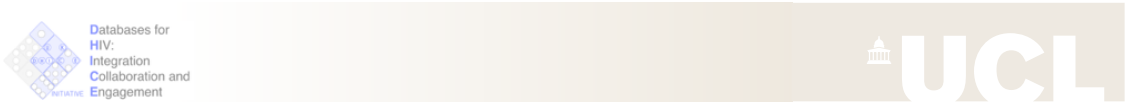
- Defines a mitigation strategy to protect resources

Databases for HIV: Integration Collaboration and Engagement

UCL

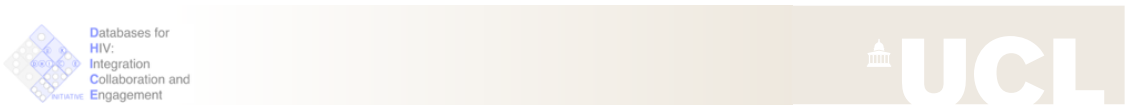
Assets,Threats,Vulnerabilities and Mitigation





Other policy details and security management

- Important to use forums within an organisation to develop policy (Information Security Management Forum)
  - user engagement
  - ensuring that they know what is in a policy and are engaged
  - management is committed
- Nothing can be guaranteed - but frequently reviewed and updated security procedures and policy management make it far less likely that information is compromised



Further thoughts - the nature of Security...

- Requirements change and evolve
  - new technology
  - more detailed information more readily available
- Other bodies decide whether information should be shared
  - Ethics Committees
  - National Information Governance Board (NIGB)

Databases for HIV: Integration Collaboration and Engagement

UCL

Further Information

- NHS
  - Connecting for Health (CfH) - http://www.connectingforhealth.nhs.uk/
  - Care Record Guarantee - http://www.nigb.nhs.uk/pubs/nhscrg.pdf
- NRES - http://www.nres.nhs.uk/
- NIGB - http://www.nigb.nhs.uk/
- ICO - http://www.ico.gov.uk/
- ISO 27000 Series

Please contact the author if you would like to see Appendix 5.

Appendix 6. Help Articles in *keibi*

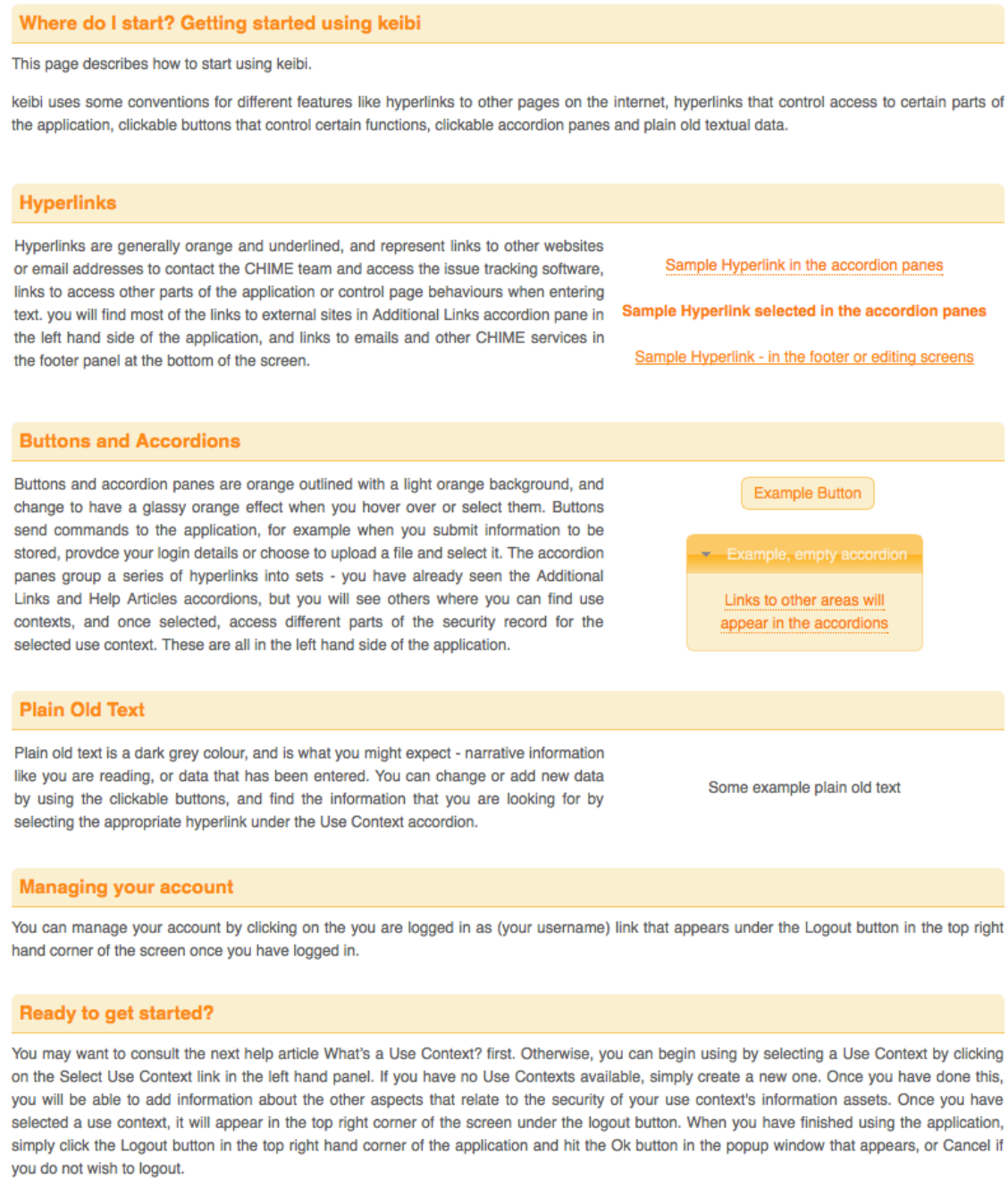


Figure 54:Getting started help ecreen

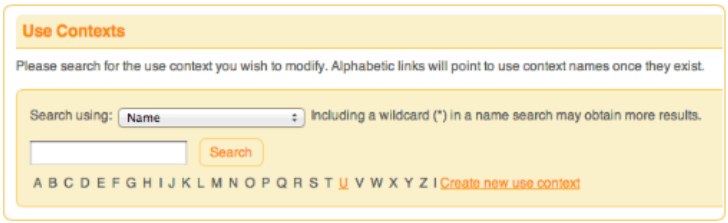
Use Contexts - what are they?

This page describes Use Contexts in keibi.

A Use Context is exactly what it says on the tin: it represents the context that you are working in as you use information for your tasks and defines the scope of the security and governance stipulations that apply to your Use Context as you work. You will add details about the information assets, activiities, their safeguards and other details to the Use Context, and be able to review these details for each Use Context since they are linked directly to it when you have created or selected the one you are interested in.

How do I choose a Use Context?

Simply click on Select Use Context under the Use Context Accordion, and use the panel that appears to either search for a Use Context by name or acronym, select from the Alphabet List or create a new Use Context where indicated.



What kinds of information should I enter about Use Contexts?

Here are a few tips to help you get started entering information about Contexts.

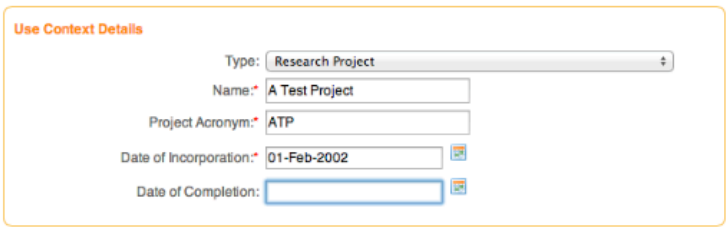
A Use Context could be a specific health research or surveillance project, or refer to a general policy for an entire organisation or department. Make sure you are sure about which one you want to add before you start - you should select which one applies to your Use Context from the Type drop down list.

Entering Use Context Main Details

Make sure that you fill in a full name and a start date - these are mandatory.

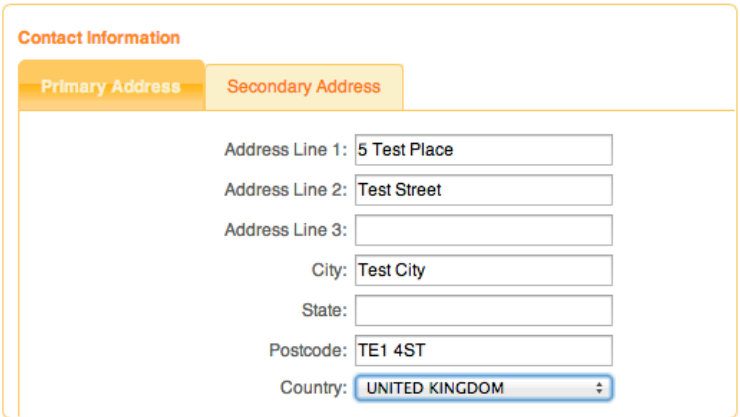
If the Use Context has ended (for example a project or study has been completed) just enter the date under completion date.

Most Use Contexts have an Acronym: if this is the case, just enter it here, and if not, simply type Not Applicable in the text input box.



Entering Addresses

You should supply for Primary Address the address of the primary co-ordinator of the Use Context, and add a secondary address if there are other correspondence addresses.



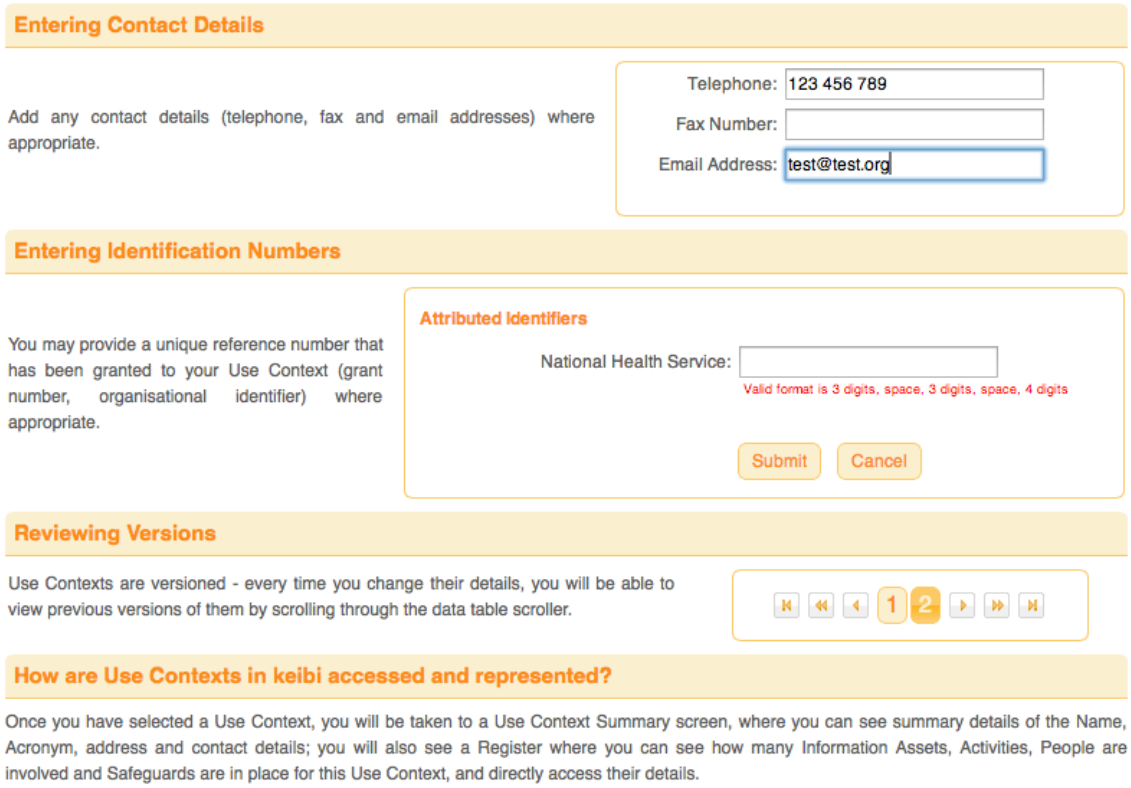


Figure 55: About Use Contexts help screen

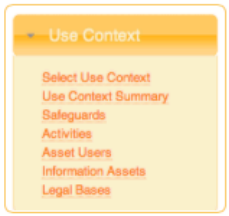
Activities - what are they?

This page describes Activitiess in keibi.

An Activity in keibi represents activities that you might perform with Information Assets. It might be that you are running a clinical governance review, an audit or presenting results of your latest study at a research conference.

How do I create an Activity?

Having selected your Use Context, simply select Activities from the left hand panel.



What kinds of information should I enter about Activities?

Here are a few tips to help you get started entering information about Activites.

A Use Context could be a specific health research or surveillance project, or refer to a general policy for an entire organisation or department. Make sure you are sure about which one you want to add before you start - you should select which one applies to your Use Context from the Type drop down list.

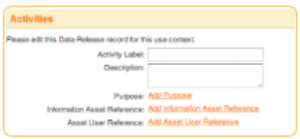
Entering Use Context Main Details

Label your Activity so it is clear what it represents

Adding a description will help other users understand what the Activity represents.

You can select a series of purposes for your activity - there can be more than one.

Feel free to add specific asset users to your Activities where appropriate.



Reviewing Versions

Activities are versioned - every time you change their details, you will be able to view previous versions of them by scrolling through the data table scroller.

Figure 56: About Activities help screen

What are Information Assets?

This page describes Information Assets in keibi.

An Information Asset in keibi represents any kind of asset that holds or provides information, be it a field in a database, the database itself, a USB key that holds data or a server that runs database and record keeping software holding a set of data.

How do I create an Information Asset?

Having selected your Use Context, simply select Information Assets from the left hand panel.

What kinds of information should I enter about Information Assets?

Here are a few tips to help you get started entering information about Activites.

Entering Informat Asset Details

Label your Information Asset so it is clear what it represents

Adding a description will help other users understand what the Information Asset represents.

You can select enter an unique identifier for your Information Asset, such as a serial number for an external hard disk.

You can add details about the type of asset that you are recording.

Some assets refer to data fields or objects of information that are stored according to a particular information model. Only use this field if it applies.

You can specify how information assets are stored where appropriate.

Reviewing Versions

Information Assets are versioned - every time you change their details, you will be able to view previous versions of them by scrolling through the data table scroller.

Figure 57: About Information Asset help screen

What are Safeguards?

This page describes how to start working with Safeguards.

Safeguards provide the details about the protection measures for information assets within your Use Context, taking into account the Activities that are being performed and the Asset Users that are involved with those Activities. The list below describes each of the fields for a Control, and the purpose for their inclusion.

How do I create an Activity?

Having selected your Use Context, simply select Activities from the left hand panel.

What kinds of information should I enter about Safeguards?

Here are a few tips to help you get started entering information about Safeguards.

Entering Safeguard Main Details

Label your Safeguard so it is clear what it represents

Adding a description will help other users understand what the Safeguard represents.

You can provide a reference to the Legal Basis for a Safeguard where appropriate.

You should refer to Activities that this Safeguard applies to, where appropriate.

You should refer to Information Assets that this Safeguard applies to, where appropriate.

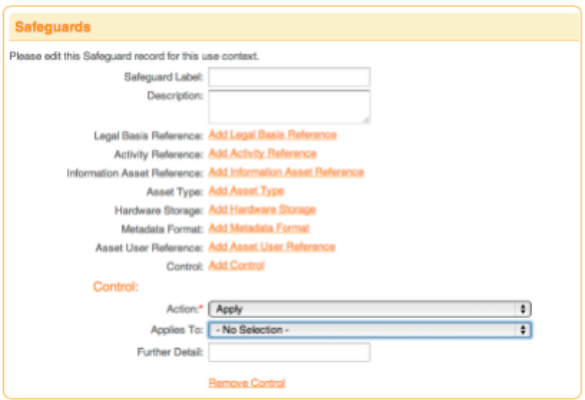
Where a Safeguard refers to a particular asset type only, select the appropriate ones here.

Where a Safeguard refers to a particular kind of hardware storage, select the appropriate ones here.

Where a Safeguard refers to a particular kind of metadata format, select the appropriate ones here.

Asset Users that this safeguard applies to should be added here.

A Control provides three fields to enter: an Action, Apply, Forbid and Permit, what that Action Applies To, be it Behaviour, Access or Release. Any Further Details should be supplied in that field. For example, if a policy denies access to patient records after 5:00pm on a Wednesday to Brian, you would specify the asset user Brian, the Information Asset Patient Records, and in teh Control, the Action Deny, the Applies to Access and the Further Detail after 5:00pm on Wednesdays..



Reviewing Versions

Safeguards are versioned - every time you change their details, you will be able to view previous versions of them by scrolling through the data table scroller.

Figure 58: About Safeguards help screen

What are Asset Users?

This page describes Information Assets in keibi.

An Asset User in keibi represents any kind of User that works with Information Assets. You can specify details about any user associated with your Use Context.

How do I create an Asset User?

Having selected your Use Context, simply select Asset Users from the left hand panel.

What kinds of information should I enter about Asset Users?

Here are a few tips to help you get started entering information about Asset Users.

Entering Asset User Details

You can provide a name and surname for your asset user.

Adding a description will help other users understand more about the Asset User.

Entering a job title is important to illustrate what the Asset User is doing within a use Context.

Asset Users may be involved with several different organisations - you can list them here.

Providing details of Asset User responsibilities helps give more of an idea about what kinds of data access they require..

Adding role details provides an indication of the kinds of groups that asset users belong to.

Reviewing Versions

Information Assets are versioned - every time you change their details, you will be able to view previous versions of them by scrolling through the data table scroller.

Figure 59: About Activities help screen

What are Legal Bases?

This page describes Legal Bases in keibi.

An Legal Basis describes the legal mechanism that allows the processing of sensitive identifiable information. It may be section 251 exemption, subject / participant consent or an employment contract. You can upload documents that provide the legal basis, or a URL if it is held online.

How do I create a Legal Basis?

Having selected your Use Context, simply select Legal Bases from the left hand panel.

What kinds of information should I enter about Legal Bases?

Here are a few tips to help you get started entering information about Legal Bases.

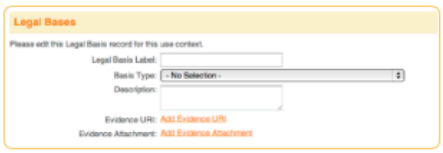
Entering Legal Basis Details

You should clearly label the Legal Basis.

Adding a description will help other users understand more about the Legal Basis.

Select the appropriate type of Legal Basis.

You may add any documents or online resources that provide the evidence of the legal basis.



Reviewing Versions

Information Assets are versioned - every time you change their details, you will be able to view previous versions of them by scrolling through the data table scroller.

Figure 60: About Legal Bases Help Screen

Appendix 7. Learning Outcomes from Pilot Evaluations

The points below summarise the learning outcomes form running the pilot evaluations. These were used to update the knowledge management framework and further develop the evaluation approach.

1. Pilot participants felt disappointed that they only got to author and use the Safeguards. The main evaluations should also test the other Secutype model usage within *keibi*.
2. The server running *keibi* required a restart during the evaluations. It would be advisable to use a dedicated infrastructure that is running its own network to mitigate the risks of software or hardware failure, and include a backup solution for the main evaluation days.
3. Pilot participants felt that the knowledge model was too detailed, particularly the inclusion of a Safeguard type: they felt that this could be removed without losing any understanding or useful detail.
4. Expected outcomes should be more fully specified in reviewing the responses. It was clear on further analysis of the results that a more detailed categorisation of errors would be needed. This should include details about the reasons for the error, be it as a result of the participant making a mistake, the tool misguiding the participant or a misunderstanding of the original excerpt. Omissions should also be recorded and subcategorised according to the reasons.
5. Examples of where participants exceeded expectations and came up with responses that had not been expected should be recorded to gather more results about the effects of the tool. The main evaluations should record where participants used their own expertise to develop responses.
6. More time should be allowed to approach participants and gather the excerpt spreadhseets. This would allow for more opportunity to analyse the provided spreadsheets and develop exercise sheets. In the pilot evaluations, little time was available to assess the excerpts and develop the exercise sheets.

Appendix 8. Confirmation of Exemption from Ethical Approval

From: GradSch.Ethics <ethics@ucl.ac.uk>
Subject: RE: Query regarding exemption from formal ethical approval
Date: 16 October 2013 13:52:38 BST
To: "Lea, Nathan" <n.lea@ucl.ac.uk>

Dear Nathan

See our exemption categories below. Your study seems to fall into the first bullet. Helen

Is my research EXEMPT from requiring ethical approval?

Within the definition of research, the following are not considered 'research' and would be exempt:

* Service evaluation undertaken to benefit those who use a particular service and is designed and conducted solely to define or judge current service. Your participants will normally be those who use the service or deliver it. It involves an intervention where there is no change to the standard service being delivered (e.g. no randomisation of service users into different groups). This does not require ethical approval.

It is possible to use data collected from participants during a service evaluation for later research as long as the data is completely anonymous; it is not possible to identify participants from any resulting report;

use of the data will not cause substantial damage and distress.

* Performance reviews
* Quality assurance/audit projects that do not involve access to or collection of private or sensitive data.
* Testing within normal education requirements

* Literary or artistic criticism

Helen Dougal
Research Ethics Co-ordinator
UCL Graduate School
North Cloisters
Wilkins Building
Gower Street
London WC1E 6BT
Tel: 020 7679 7844 (ext 37844)
Email: ethics@ucl.ac.uk

Working hours: Mon-Fri 8am-3.30pm

-----Original Message-----
From: Lea, Nathan
Sent: 13 October 2013 20:45
To: GradSch.Ethics
Subject: Query regarding exemption from formal ethical approval

I have a query regarding a set of evaluations I am planning to start for my PhD research that will involve human participants and I wondered if I could seek your advice on whether the research will be exempt from formal ethical approval. I have attached a Microsoft Word document that summarises the planned evaluations, which provides details of how the participants will be involved and how I am planning to conduct the research. I am not sure whether formal ethical approval is needed because information obtained from the research work will not be recorded in a way that will identify the participants, but I think it would be best to confirm this is the case with you before I proceed.

Please let me know if you have any further questions about this and many thanks in advance for your assistance.

With best wishes,

Nathan

Nathan C. Lea
Senior Research Associate
http://www.ucl.ac.uk/chime/people/lean
Tel: +44(0) 20 3549 5293
Fax: +44(0) 20 7679 5064

CHIME HAS MOVED! You can now find us at: 3rd Floor, Wolfson House, 4 Stephenson Way, London NW1 2HE.
For more information about studying at CHIME, see http://www.ucl.ac.uk/chime/study

Appendix 9. Invitation Email and Introduction Document

The invitation email and introductory document are below:

From: Nathan Lea <n.lea@ucl.ac.uk>
Subject: Invitation to participate in PhD evaluations
Date: 31 October 2013 14:47:25 GMT
To: xxxxx

Dear xxx,

You may recall that I have been preparing to evaluate a software tool I have developed as part of my PhD work. Part of the evaluation involves people using the tool in a laboratory environment, and I am delighted to invite you to participate.

I have attached a PDF that gives a broad overview of what will be involved and the contributions that you will be asked to make. Could I ask you to read this over and let me know if you would like to participate? I would be very grateful if you could get back to me in the next few days - I am hoping to hold the sessions in late November through to mid December 2013.

Do please let me know if you have any questions about this at all.

With best wishes,

Nathan

Nathan C. Lea
Senior Research Associate
http://www.ucl.ac.uk/chime/people/lean
Tel: +44(0) 20 3549 5293

Fax: +44(0) 20 7679 5064

CHIME HAS MOVED! You can now find us at: 3rd Floor, Wolfson House, 4 Stephenson Way, London NW1 2HE.

For more information about studying at CHIME, see

http://www.ucl.ac.uk/chime/study

Evaluation_Introductory_Document.pdf:

Invitation to Participate in the Evaluation of an Information Security Policy Software Tool

I am writing to invite you to participate in a series of studies about using a new software tool to manage information security as part of your working practice. I am approaching you because you work with sensitive healthcare information, or you are responsible for establishing good working practice and policy for its safe and appropriate processing. You may fall into both of these categories, and your working practice may be governed by these policies and you have to apply them in practice.

The tool that has been developed as part of my PhD thesis on the protection of privacy when information captured during routine care purposes is reused for wider purposes, including research, public health policy development and medical education. The software is a web based tool called *keibi* that is designed to make the process of managing security and understanding how you should behave with healthcare information clearer and easier, whilst reducing the need to remember numerous policy stipulations and decide how best to apply them for each situation that arises.

To validate these claims, the studies have been designed to evaluate the use of the tool in a laboratory setting using fictitious information: no actual patient records will be used

or accessed. Your participation in the studies will involve you in the following ways:

1. You will be asked to supply some examples of information security policy, data sharing agreement and working practice examples that you use within your organisation (see notes in next paragraph).

2. You will be asked to attend a three hour workshop along with up to four other participants, where you will be asked to author the policy examples in *keibi*; you will also use policy details held by *keibi* to answer a set of questions about how you would behave with information. You will be asked to complete a short questionnaire about your experience and satisfaction using *keibi*. The workshop will conclude with a round up session where we will discuss your views together in a group;

3. You will be asked to confirm that a series of data release decisions have been made by an electronic healthcare record system in accordance with the policies you specified.

The policy examples you will be asked to provide will include three excerpts from an organisational policy document that specify how information assets should be used and protected, three from any data sharing agreements that you have with other organisations

and three examples of common working practice that involves the protection of information you process that is not to your knowledge written in any policy document. If your policies are not made publicly available and you are concerned about releasing specific details, you will be able to remove any sensitive or identifying details and replace them with generic details. You will receive guidelines to

make the task as easy and quick as possible.

The workshop will last no longer than three hours and lunch will be provided. This email summarises the main elements in the studies, and more details will of course be supplied if you agree to participate. Note that the information collected from the studies will NOT identify you or be used as a means to judge your performance on a professional level. You will be assigned a participant number and remain anonymous during the process of analysing the data and in any published proceedings of the experiments, where the only references made to you will be about your position and role within your organisation, which will also remain anonymous and referred to as only “a hospital / research institution in a UK inner city area”. You will of course be able to withdraw from the studies at any point without providing any explanation if you so wish.

I would be grateful if you could let me know if you would like to participate as soon as you are able. Once you have confirmed your participation, I will contact you with further details. Please feel free to contact me with any questions that you may have about this.

Nathan Lea (n.lea@ucl.ac.uk)

Appendix 10. Excerpt submission email and guidelines

From: Nathan Lea <n.lea@ucl.ac.uk>

Subject: Re: Invitation to participate in PhD evaluations

Date: 6 November 2013 15:20:16 GMT

To: xxxx

Dear xxxx,

Many thanks again for agreeing to take part. Further to the email that I sent you yesterday regarding your availability, I am pleased to attach further details about the evaluations.

You will recall that the first part of the evaluation was to provide me with some example policy excerpts. I have attached an Excel Spreadsheet (Participant_Policy_Excerpts_2013.xls) where you can enter the details. I have also provided a set of guidelines on how to complete the spreadsheet, which you will find in the attached PDF (Excerpt_Submission_Guidelines.pdf).

I would be grateful if you could return the completed spreadsheet by Friday 15th November 2013 if at all possible. Please do let me know if you have any questions about this.

Do please let me know which of the dates would suit you as soon as you are able. The dates I suggested yesterday are as follows:

Friday 22nd November 10:00am - 1:00pm
Wednesday 27th November 10:00am - 1:00pm
Monday 2nd December 2:00pm - 5:00pm
Tuesday 3rd December 10:00am - 1:00pm
Wednesday 11th December 10:00am - 1:00pm.

Do please get in touch if you have any further questions.

With best wishes,

Nathan

Nathan C. Lea
Senior Research Associate
http://www.ucl.ac.uk/chime/people/lean
Tel: +44(0) 20 3549 5293
Fax: +44(0) 20 7679 5064

CHIME HAS MOVED! You can now find us at: 3rd Floor, Wolfson House, 4 Stephenson Way, London NW1 2HE.
For more information about studying at CHIME, see
http://www.ucl.ac.uk/chime/study

Excerpt_Submission_Guidelines.pdf

Guidelines for submitting policy excerpts for use in the evaluations

You will recall that we will be using examples of actual information security management details to evaluate *keibi*. These details are often found in information security policies, guideline documents and data sharing agreements, but they sometimes exist as good practice and are passed on by word of mouth or formed through experience but are not necessarily written in available documentation. I am asking you to provide the following in the supplied spreadsheet:

4. up to 3 excerpts from an organisational policy document that is used by you and your colleagues to

ensure safe working practice is adhered to when handling healthcare information;

5. up to 3 excerpts from a data sharing agreement that your organisation may have entered into with another organisation, where agreed protection measures are listed in this document;

6. up to 3 examples of good working practice that you use or have been told in training sessions or seminars, which are not to your knowledge documented anywhere in your organisation in a policy or data sharing agreement.

Please see the Excel Spreadsheet called Participant_Policy_Excerpts_2013.xls attached to the covering email, where you can enter the details. To help you with this, please refer to the notes below.

Notes on Completing the Spreadsheet

You should refer to policies, data sharing agreements and working practice that is already in the public domain (either provided online or otherwise already shared publicly). In order to fully test the tool, I would be grateful if you could include the following from your documents and working practice:

1. Examples of control stipulations that are intended to guide people's behaviour (for example stipulations on the use of USB keys, or how to handle printed documents) as well as those that are intended for electronic system and computer use and or configuration, like providing access to certain record items according to legitimate relationships or removal of identifying attributes for research use; under Column B in the spreadsheet labelled "**Excerpt**" place the excerpt that you have selected;

under Column C labelled "**Source**" place the source of the excerpt, be it a policy document, data sharing agreement, or a specification of good working practice that has not otherwise been documented. If your policy / data sharing agreement document is available online, or this comes from working practice that has been detailed online, please specify the URL under column D "**Online Location**."

For each excerpt, please give a concise description of what you expect the outcome of the policy excerpt to be under column E "**Intended Outcome**" and specify whether it is is for human readership, computer configuration, or both under column F: for example, you might specify that a member of staff should never use USB keys to transfer or share identifiable medical records under any circumstances, where the outcome would be that staff members do not put identifiable data on USB keys; a software configuration example might be to specify that dates of birth should not be shared for a specific secondary use like a Clinical Governance Meeting, where the outcome would be that the software does not release dates of birth when data is exported for this purpose.

Please remember that the policies and working practice should not refer to healthcare information itself, and what you provide should neither include patient records themselves or identify any patient. You should remove the name of the organisation and any staff members, rendering the policy details, organisation and staff anonymous, perhaps replacing them with fictitious names (like the Alpha Centauri Healthcare Trust or Dr. James T. Kirk).

Please avoid providing excerpts from general guideline documents like the Care Record Guarantee, the Department of Health Guidelines on information security policies, or examples from the Information Commissioner's Office. If your information security policies happen to be

available for public review, please point me to the web address of the document. If you cannot find appropriate documentation, please supply a brief written summary of no more than nine examples of working practice that is outlined in the third item in the list above (examples of good working practice that you use or have been told in training sessions or seminars, which are not to your knowledge documented anywhere in your organisation).

The reason that you are being asked to provide the materials for the evaluation is to ensure a more realistic test of the *keibi* based upon actual working practice; whilst I can supply state-of-the-art policy excerpts in a more fictitious scenario, this might skew the effects of using the tool and call into question the reliability of some of the results. Please contact me (n.lea@ucl.ac.uk) if you have any questions about this document.

POLICY EXCERPT DETAILS SPREADSHEET	Excerpt	Source (Policy Document, Data Sharing Agreement or Working Practice	Online Location	Intended Outcome (please also specify whether these are for human readership, software configuration, or both
1				
2				
3				
4				
5				
6				
7				
8				
9				

Table 36: Excerpt Submission Spreadsheet

Please contact the author if you would like to see Appendix 11.

Please contact the author if you would like to see Appendix 12.

Appendix 13. Evaluation Session Exercise Sheets

Exercise sheet 1

This document contains details you will need for the experiments that we will be working on this morning. The first experiment involves policy authoring using the excerpts that you have provided Nathan over the last few weeks. The second involves you answering a series of questions based on the policy items that have been authored by others. Please feel free to ask any question that you may have; Nathan will answer them to the best of his ability, and will make a note of the question that you asked, when you asked it and details of his response for the purposes of analysing the results. The first section asks a few questions about your work and professional background.

Some Details About You:

Please briefly describe the work that you do with sensitive information and / or your involvement in developing information governance good practice and security policies:

Please provide a brief overview of your professional background (including any degrees and / or other professional qualifications).

First Experiment: Authoring Policy Items in *keibi*

This experiment will use the policy excerpts below that you will be asked to author in *keibi*. Please author them to the best of your ability and feel free to ask any questions you like during the authoring step. Nathan will record that you asked a question, when you asked it and what you asked for analysis later on. The first three questions are a warmup designed to get you used to the process of authoring excerpts in *keibi*. Question 4 will be timed and will be the same across all of the participants, so that will be a little more formal. You will then be asked to complete questions 4, 5, 6 and 7 on your own, in your own time (maximum time allowed: ten minutes); I will time these as well, but not use them to compare across participants.

Log in to *keibi* as papdtwozero with password passwd00020 under the account Evaluation 22nd January 2014 as Clinical Care.

Select Use Context Research Project 010.

Question 1: Warmup question for everyone to answer

Please enter the information asset "CD ROM" into *keibi*, completing the details as you see fit.

Question 2: Warmup question for everyone to answer

Please enter the Legal Basis "Consent" into *keibi*, completing the details as you see fit.

Question 3: Warmup excerpt for everyone to answer

Please enter the following excerpt into *keibi*:

"**Documents with patient data even if anonymised is not to be sent by email or posted on a CD.**"

Question 4: Timed excerpt for everyone to answer

Please enter the following excerpt into *keibi*:

"**Key identifying fields, such as name, address, full postcode, NHS**

Number, will not be extracted for use in a research project"

Question 5: Excerpt for you to enter:

Please enter the following excerpt into *keibi*, ensuring that you enter the appropriate Activity and Asset Users. N.B. you should add one Asset User with the name Gustavo Fring and other details as appropriate:

"**Audit: The Processor will permit the UCL School of Life and Medical Sciences (SLMS) to monitor compliance with the terms of this agreement, which may involve the UCL SLMS or its nominated representative coming onto any premises where the personal data are being processed with at least 10 working days' notice.**"

Question 6: Excerpt for you to enter:

Please enter the following excerpt into *keibi*:

"**The use of known telephone numbers is a requirement. Under no circumstances should personally identifiable information be given to an individual calling from an unknown telephone number.**"

Question 7: Excerpt for you to enter:

Please enter the following excerpt into *keibi*, including any appropriate information assets:

"**In the event that unauthorised access was gained through security codes or passes being compromised the Line Manager must immediately take appropriate action to prevent further unauthorised access. Security codes should be changed and any passes used for unauthorised access should be disabled.**"

THE SECOND EXPERIMENT STARTS BELOW.

Second Experiment: using *keibi* to answer how you would proceed with sensitive information

The following questions are designed to see how you would respond to a series of data handling challenges based upon details that are provided to you by *keibi*.

Select the Research Project 011 Use Context.

Consult the policy details within *keibi* and answer the following questions:

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Consult *keibi* and briefly specify below how you would respond to the request: make sure that you include whether you would agree or not to the request and details of how you might handle the requested information:

Question 2: You have been asked to transfer data from a Mac Mini holding a research data repository to a colleague using a USB key. How would you respond?

Consult *keibi* and briefly specify below how you would respond to the request. Make sure that you indicate the measures that you will take prior to sharing the information, what policies in *keibi* you are following and justify your response:

Question 3: A colleague has requested that you network the Mac Mini to perform a critical software update. How would you respond?

Consult *keibi* and briefly specify below how you would respond to the request. You may assume that the request is genuine and reasonable, subject to the policies specified in *keibi*.

My Choice		Response Details
	1	I would network the Mac Mini to perform the requested function.
	2	I would network the Mac Mini and have the requesting party sign a disclaimer that the policy was breached to perform an essential function.
	3	I would decline the request, citing the information security policy.
	4	I would speak to the policy authors and ask them whether this is an acceptable case for breaching the policy, or how they would proceed.

Please briefly justify your response below:

Question 4: A research colleague has asked for some patient data. They say that it can be anonymised and that they would be able to receive it via email, on a posted CD or they will be happy to collect it on an encrypted USB key. How do you respond?

Consult *keibi* and briefly specify below how you would respond to the request. You may assume that the request is genuine and reasonable, subject to the policies specified in *keibi*.

Exercise sheet 2

This document contains details you will need for the experiments that we will be working on this morning. The first experiment involves policy authoring using the excerpts that you have provided Nathan over the last few weeks. The second involves you answering a series of questions based on the policy items that have been authored by others. Please feel free to ask any question that you may have; Nathan will answer them to the best of his ability, and will make a note of the question that you asked, when you asked it and details of his response for the purposes of analysing the results. The first section asks a few questions about your work and professional background.

Some Details About You:

Please briefly describe the work that you do with sensitive information and / or your involvement in developing information governance good practice and security policies:

Please provide a brief overview of your professional background (including any degrees and / or other professional qualifications).

First Experiment: Authoring Policy Items in *keibi*

This experiment will use the policy excerpts below that you will be asked to author in *keibi*. Please author them to the best of your ability and feel free to ask any questions you like during the authoring step. Nathan will record that you asked a question, when you asked it and what you asked for analysis later on. The first three questions are a warmup designed to get you used to the process of authoring excerpts in *keibi*. Question 4 will be timed and will be the same across all of the participants, so that will be a little more formal. You will then be asked to complete questions 4, 5, 6 and 7 on your own, in your own time (maximum time allowed: ten minutes); I will time these as well, but not use them to compare across participants.

Log in to *keibi* as papdtwoeight with password passwd00028 under the account Evaluation 22nd January 2014 as Clinical Care.

Select Use Context Research Project 011.

Question 1: Warmup question for everyone to answer

Please enter the information asset "CD ROM" into *keibi*, completing the details as you see fit.

Question 2: Warmup question for everyone to answer

Please enter the Legal Basis "Consent" into *keibi*, completing the details as you see fit.

Question 3: Warmup excerpt for everyone to answer

Please enter the following excerpt into *keibi*:

"**Documents with patient data even if anonymised is not to be sent by email or posted on a CD.**"

Question 4: <u>Timed</u> excerpt for everyone to answer

Please enter the following excerpt into *keibi*:

"**Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project**"

Question 5: Excerpt for you to enter:

Please enter the following excerpt into *keibi*, ensuring that you enter the appropriate Activity and Asset Users. N.B. you should add one Asset User with the name Walter White and other details as appropriate:

"**All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.**"

Question 6: Excerpt for you to enter:

Please enter the following excerpt into *keibi*:

"**The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances.**"

Question 7: Excerpt for you to enter:

Please enter the following excerpt into *keibi*:

"**No copies of data shall be made or distributed from the Mac Mini on any removable medium.**"

THE SECOND EXPERIMENT STARTS BELOW.

Second Experiment: using *keibi* to answer how you would proceed with sensitive information

The following questions are designed to see how you would respond to a series of data handling challenges based upon details that are provided to you by *keibi*.

Select the Research Project 010 Use Context.

Consult the policy details within *keibi* and answer the following questions:

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Consult *keibi* and briefly specify below how you would respond to the request: make sure that you include whether you would agree or not to the request and details of how you might handle the requested information:

Question 2: You have received a telephone request for names and addresses of three participants in a research trial. How would you respond?

Consult *keibi* and briefly specify below how you would respond to the request. Make sure that you indicate the measures that you will take leading up to your decision about sharing sharing the information and justify your response:

Question 3: A team member from UCL SLMS has asked to come and inspect your compliance with a data sharing agreement you hold with them within the next week. How would you respond to their request?

Consult *keibi* and briefly specify below how you would respond to the request. You may assume that the request is genuine and reasonable, subject to the policies specified in *keibi*.

My Choice		Response Details
	1	I would agree and offer dates.
	2	I would not agree to this request and refer the matter to my line manager.
	3	I would agree to the request, but only allow access over the telephone.
	4	I would agree and offer dates ten or more days later.

Please briefly justify your response below:

Question 4: You have received a report of an unauthorised access to an information asset. How do you respond to this?

Consult *keibi* and briefly specify below how you would respond to the request. You may assume that the report is genuine and reasonable, and your response must be subject to the policies specified in *keibi*.

Appendix 14. Description of Exercise Sheets and Expected Answers

Experiment One - Authoring Policy

There were a total of seven questions in each of the exercise sheets, where the first four were the same across each sheet. The final three were based upon provided policy excerpts, making a total of ten different questions in the first experiment. These are listed below with their expected answers.

Question 1: Warmup question for everyone to answer

Please enter the information asset "CD ROM" into *keibi*, completing the details as you see fit.

This question was provided as a warm up and a means to add an Information Asset that would be reused in other questions. It was also added in response to the pilot cohort expressing a desire to have used more of *keibi*'s features other than authoring Safeguards only. Participants were expected to enter a new Information Asset of CD ROM. They could add further details as they saw fit.

Question 2: Warmup question for everyone to answer

Please enter the Legal Basis "Consent" into *keibi*, completing the details as you see fit.

As with Question 1, this question was provided to give participants an opportunity to try out a different aspect of *keibi* other than the Safeguards, as well as provide a Legal Basis that would be reused in other questions. It expected participants to enter a Legal Basis of consent, where they could enter other details as they saw fit.

Both questions one and two were provided so that participants could get a feel for when they should be entering additional details as they read through the policy excerpts.

Question 3: Warmup excerpt for everyone to answer

Please enter the following excerpt into *keibi*:

"Documents with patient data even if anonymised are not to be sent by email or posted on a CD."

This question was provided to give participants some practice in authoring a Safeguard. It deliberately included reference to the CD ROM Information Asset to see if participants could include that in the Safeguard itself, though the correct course of action would be to specify the Hardware Storage of CD ROM. There was also an opportunity to add an Activity to cover sharing data. It had more expectations: it expected a Safeguard to be authored, as well as an Activity that involved the sharing of information. There are no specific expectations on how this is expressed in the Safeguard, but at the very least, the Safeguard should refer to the sharing information Activity and CD ROM Information Asset, where the requisite Control should refer to anonymisation and make clear that sharing information on CD or via email should be forbidden.

Question 4: Timed excerpt for everyone to answer

Please enter the following excerpt into *keibi*:

"Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project"

This question was supplied to ensure that at least one set of computable policy items were available. It expected the participant to author a Safeguard that would provide some heuristics for software configuration, including a control that forbade the release of these fields. By way of preparation, the four fields that needed to be referred to were added by the investigator prior to the evaluation sessions so that participants only had to focus on the Safeguard that needed to be prepared. This Question also expected participants to author an Activity that related to research or running research projects.

Question 5: Excerpt for you to enter:

Please enter the following excerpt into *keibi*, ensuring that you enter the appropriate Activity and Asset Users. N.B. you should add one Asset User with the name Walter White and other details as appropriate:

"All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing."

This question is harder to answer as it requires the participants to look for the specific *keibi* centred detail in general statements. The expected response includes at least one Safeguard that specified researchers must obtain consent before sharing data for a series of Activities, including research, cohort studies and clinical trials (though it was left to the participant to decide how they specified these Activities, whether it was one Activity that covered the three examples, or three separate ones for each). This Safeguard should therefore include the sharing data Activity, research, cohort studies and clinical trials Activities (if they already existed) and specify that data should not be shared without consent under the

Control in the Safeguard. Participants were asked to specify a researcher, in this case "Walter White" so that they could apply an Asset User to the Safeguard.

Question 6: Excerpt for you to enter:

Please enter the following excerpt into *keibi*:

"The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances."

This question provided a specific set of Information Assets and technical details allowing participants to prepare a set of Safeguards and other details that focused on those Assets. The expected response was for participants to create a Mac Mini Information Asset, a PDA Information Asset, and a Safeguard that prevents the Mac Mini from being networked or connected to another device in any of the listed ways (ethernet, Bluetooth etc.). One way this could be achieved is to have a Control specified for each of the connection methods.

Question 7: Excerpt for you to enter:

Please enter the following excerpt into *keibi*:

"No copies of data shall be made or distributed from the Mac Mini on any removable medium."

This question was used to allow participants to carry forward some of the policy items that they had, in theory, already authored, and apply a single Safeguard to prevent the copying of data to removable media. They had the option of either creating a new Information Asset to cover removable media, or using the options

under the Controls in the expected Safeguard to make this clear. The interpretation of this excerpt would be key in discovering how they specified this.

The following three questions were used in the second answer sheet set. Three different questions were used from the either set so that in exercise two, the counterpart participants in the session would have an opportunity to answer questions using excerpts that they had not authored in *keibi* for experiment two, and vice versa. The third set of exercise sheets remained the same as this, merely replacing the Asset User "Jesse Pinkman" with another fictional name, "Gustavo Fring."

Question 5: Excerpt for you to enter:

Please enter the following excerpt into *keibi*, ensuring that you enter the appropriate Activity and Asset Users. N.B. you should add one Asset User with the name Jesse Pinkman / Gustavo Fring and other details as appropriate:

"Audit: The Processor will permit the UCL School of Life and Medical Sciences (SLMS) to monitor compliance with the terms of this agreement, which may involve the UCL SLMS or its nominated representative coming onto any premises where the personal data are being processed with at least 10 working days' notice."

As with Question 5 of the first set of exercise sheets, this question is harder to answer requiring the participants to look for the specific *keibi* centred detail in general statements. The expected response includes an Activity called Audit, the Asset User assigned the role Auditor as the nominated representative of SLMS. There was an additional Information Asset that could be entered as Personal Data, and at least one Safeguard that specified that Pinkman / Fring would be allowed access to where the personal data with ten working day's notice. This Safeguard should therefore include the audit Activity, personal data Information Asset and at

least one Control that permitted access to the project within ten working days' notice. Participants were expected to add the Asset User Pinkman / Fring as the SLMS representative in the Safeguard.

Question 6: Excerpt for you to enter:

Please enter the following excerpt into *keibi*:

"The use of known telephone numbers is a requirement. Under no circumstances should personally identifiable information be given to an individual calling from an unknown telephone number."

As with question seven in the first set of exercise sheets, this focussed on a particular use case with Information Assets, as well as a specific Safeguard relating to its use. The Information Assets were the telephone, potentially telephone numbers that were known and personally identifiable information. The expected Safeguard should specify that the personally identifiable information should not be released to unidentified phone numbers.

Question 7: Excerpt for you to enter:

Please enter the following excerpt into *keibi*, including any appropriate information assets:

"In the event that unauthorised access was gained through security codes or passes being compromised the Line Manager must immediately take appropriate action to prevent further unauthorised access. Security codes should be changed and any passes used for unauthorised access should be disabled."

This question was similar to question seven in the first series: it focusses on a particular set of Information Assets and at least one Safeguard with multiple

Controls. The expected response was that participants created security passes and access codes Information Assets, an activity for handling unauthorised access incidents, along with a Safeguard that handled unauthorised access, with Controls that at the very least required the line manager be informed of the breach, another Control to change codes and another to disable passes.

Experiment 2 – Handling Information Assets using *keibi*

There were a total of four questions for each exercise sheet, which relied on the use of the policies authored as part of the first experiment to answer them. These questions related to the scenarios that were covered by those policy items used for the first experiment.

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spread sheet.

Consult *keibi* and briefly specify below how you would respond to the request: make sure that you include whether you would agree or not to the request and details of how you might handle the requested information:

This question was posed to test whether participants could appreciate the details of the original excerpts when presented in *keibi* based on the detail entered in questions three and four. The expected response was that the anonymised data could be shared, but not via email.

Question 2: You have been asked to transfer data from a Mac Mini holding a research data repository to a colleague using a USB key. How would you respond?

Consult *keibi* and briefly specify below how you would respond to the request. Make sure that you indicate the measures that you will take prior to sharing the information, what policies in *keibi* you are following and justify your response:

This question sought to test how well the detail of the excerpts entered under questions six and seven from the first set of exercise sheets were interpreted by participants. The expected response was that the request be denied and that how the request might otherwise be handled was not clear from the details in *keibi*.

Question 3: A colleague has requested that you network the Mac Mini to perform a critical software update. How would you respond?

Consult *keibi* and briefly specify below how you would respond to the request. You may assume that the request is genuine and reasonable, subject to the policies specified in *keibi*.

My Choice		Response Details
	1	I would network the Mac Mini to perform the requested function.
	2	I would network the Mac Mini and have the requesting party sign a disclaimer that the policy was breached to perform an essential function.
	3	I would decline the request, citing the information security policy.
	4	I would speak to the policy authors and ask them whether this is an acceptable case for breaching the policy, or how they would proceed.

This question sought to test how well the detail of the excerpts entered under question six from the first set of exercise sheets were interpreted by participants. The expected response was that participants selected item four; whilst item 3 would be correct in terms of what details would be available on *keibi*, wider good practice requires that a change or an exception to policy be considered and further advice sought as there is a conflict in not networking the Mac Mini and not

updating it. This was designed to test whether using *keibi* undermined such good practice.

Question 4: A research colleague has asked for some patient data. They say that it can be anonymised and that they would be able to receive it via email, on a posted CD or they will be happy to collect it on an encrypted USB key. How do you respond?

Consult *keibi* and briefly specify below how you would respond to the request. You may assume that the request is genuine and reasonable, subject to the policies specified in *keibi*.

This question sought to test how well the detail of the excerpts entered under questions three, four, five, six and seven would be interpreted to answer it. The expected response was that the request would be refused, but left it open as to whether alternative methods for transferring data be considered, and whether the request came from a colleague that had rights of access in the first place.

The following are the questions for experiment two from the second set of exercise sheets. The first question was the same fro both exercise sheets.:

Question 2: You have received a telephone request for names and addresses of three participants in a research trial. How would you respond?

Consult *keibi* and briefly specify below how you would respond to the request. Make sure that you indicate the measures that you will take leading up to your decision about sharing the information and justify your response:

This question expected participants to be able to use the excerpts entered in questions five and six from the second set of questions in experiment one. The expected response was that the number of the caller should be checked and verified, after which the list could be provided to the caller. Participants were also

expected to say that the details in keibi lacked some core good practice guidelines, like verifying the identity of the caller and their permission to have the information shared.

Question 3: A team member from UCL SLMS has asked to come and inspect your compliance with a data sharing agreement you hold with them within the next week. How would you respond to their request?

Consult *keibi* and briefly specify below how you would respond to the request. You may assume that the request is genuine and reasonable, subject to the policies specified in *keibi*.

My Choice		Response Details
	1	I would agree and offer dates.
	2	I would not agree to this request and refer the matter to my line manager.
	3	I would agree to the request, but only allow access over the telephone.
	4	I would agree and offer dates ten or more days later.

Please briefly justify your response below:

This question expected participants to choose answer four. Arguable answer two would also be possible, where if there was any uncertainty about the stipulations in *keibi*, participants may feel more comfortable referring to the line manager to decide, particularly since there would have been a reference to the line manager after details from question 7 in the second set of answer sheets were entered. This however would not be correct as participants were told to treat the request as valid and to respond according to the specified policy in *keibi*.

Question 4: You have received a report of an unauthorised access to an information asset. How do you respond to this?

Consult *keibi* and briefly specify below how you would respond to the request. You may assume that the report is genuine and reasonable, and your response must be subject to the policies specified in *keibi*.

This question expected participants to review the details entered in question seven of the second set of exercise sheets. The expected response was to notify the line manager, disable compromised passed and reissue security codes.

Appendix 15. Evaluation sessions introductory slides

UCL

***keibi* test evaluations**

Nathan Lea
Thursday 16th January 2014

UCL

The Afternoon

- Introduction
- Experiment 1: authoring excerpts
- Refreshment / comfort break
- Experiment 2: using *keibi* to answer data handling questions and satisfaction questionnaires
- Refreshments and Group discussion

2

UCL

Introducing *keibi*

警備 keibi

- Why 'keibi'?
- What's *keibi* for?

3

UCL

Quick Overview

- Contexts of use
- Activities
- Information Assets
- Safeguards
- Asset Users
- Legal Bases

警備 keibi

4

UCL

Any questions?

GOOD LUCK!!

5

Appendix 16. User Satisfaction Questionnaire

The Post-Study System Usability Questionnaire (PSSUQ)

Instructions and Items. The questionnaire's instructions and items are:

This questionnaire, which starts on the following page, gives you an opportunity to tell us your reactions to the system you used. Your responses will help us understand what aspects of the system you are particularly concerned about and the aspects that satisfy you.

To as great a degree as possible, think about all the tasks that you have done with the system while you answer these questions.

Please read each statement and indicate how strongly you agree or disagree with the statement by circling a number on the scale. If a statement does not apply to you, circle N/A.

Please write comments to elaborate on your answers.

After you have completed this questionnaire, I'll go over your answers with you to make sure I understand all of your responses.

Thank you!

1. Overall, I am satisfied with how easy it is to use this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

2. It was simple to use this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

3. I could effectively complete the tasks and scenarios using this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

4. I was able to complete the tasks and scenarios quickly using this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

5. I was able to efficiently complete the tasks and scenarios using this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

6. I felt comfortable using this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

7. It was easy to learn to use this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

8. I believe I could become productive quickly using this system.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

9. The system gave error messages that clearly told me how to fix problems.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

10. Whenever I made a mistake using the system, I could recover easily and quickly.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

11. The information (such as on-line help, on-screen messages and other documentation) provided with this system was clear.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

12. It was easy to find the information I needed.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

13. The information provided for the system was easy to understand.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

14. The information was effective in helping me complete the tasks and scenarios.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

15. The organization of information on the system screens was clear.

STRONGLY AGREE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**STRONGLY DISAGREE**	N/A

COMMENTS:

Note: *The interface includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the screens (including their use of graphics and language).*

16. The interface of this system was pleasant.

STRONGLY AGREE **1** **2** **3** **4** **5** **6** **7** **STRONGLY DISAGREE** N/A

COMMENTS:

17. I liked using the interface of this system.

STRONGLY AGREE **1** **2** **3** **4** **5** **6** **7** **STRONGLY DISAGREE** N/A

COMMENTS:

18. This system has all the functions and capabilities I expect it to have.

STRONGLY AGREE **1** **2** **3** **4** **5** **6** **7** **STRONGLY DISAGREE** N/A

COMMENTS:

19. Overall, I am satisfied with this system.

STRONGLY AGREE **1** **2** **3** **4** **5** **6** **7** **STRONGLY DISAGREE** N/A

COMMENTS:

Appendix 17. Exercise Results and Analysis

2nd December 2013

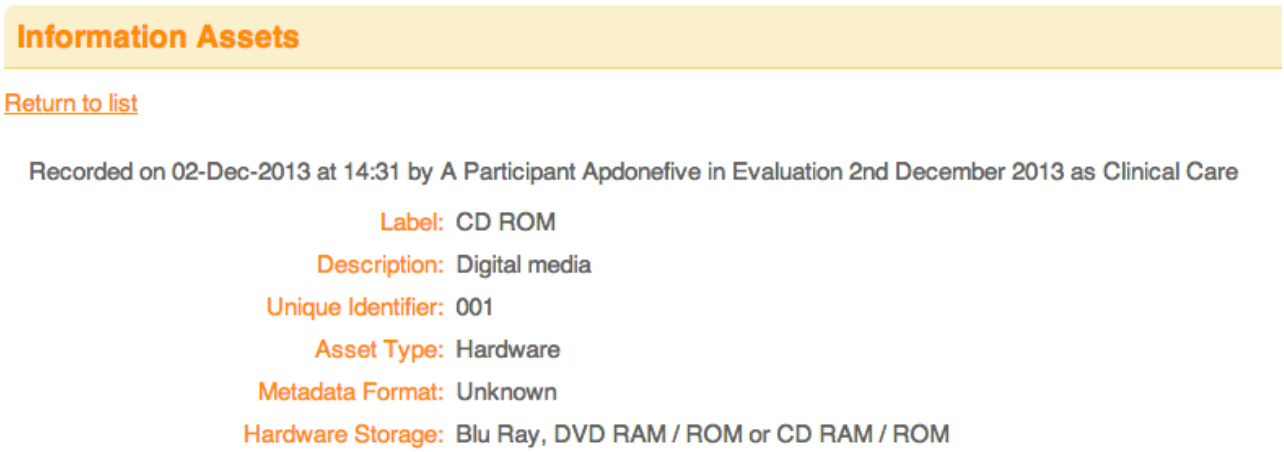
Participant APD_00015

The responses from participant APD_00015 are provided below.

Experiment One

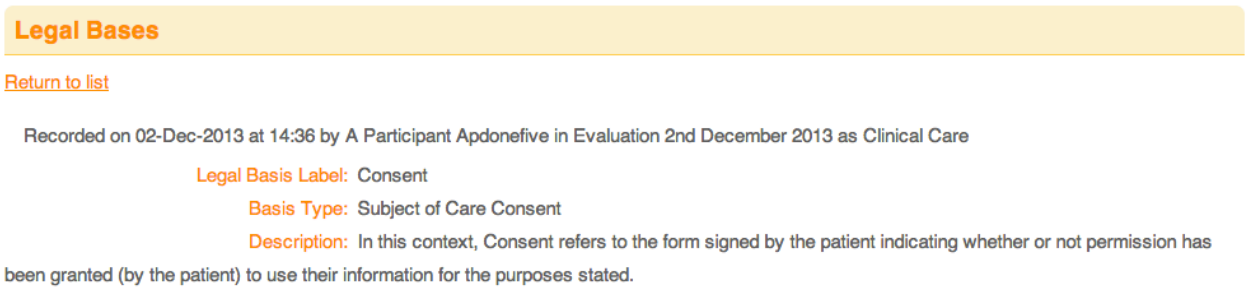
The first experiment asked all participants to author policy excerpts in *keibi*. Section xxx provides details of the questions posed in this experiment, and the excerpts are repeated here for convenience.

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.



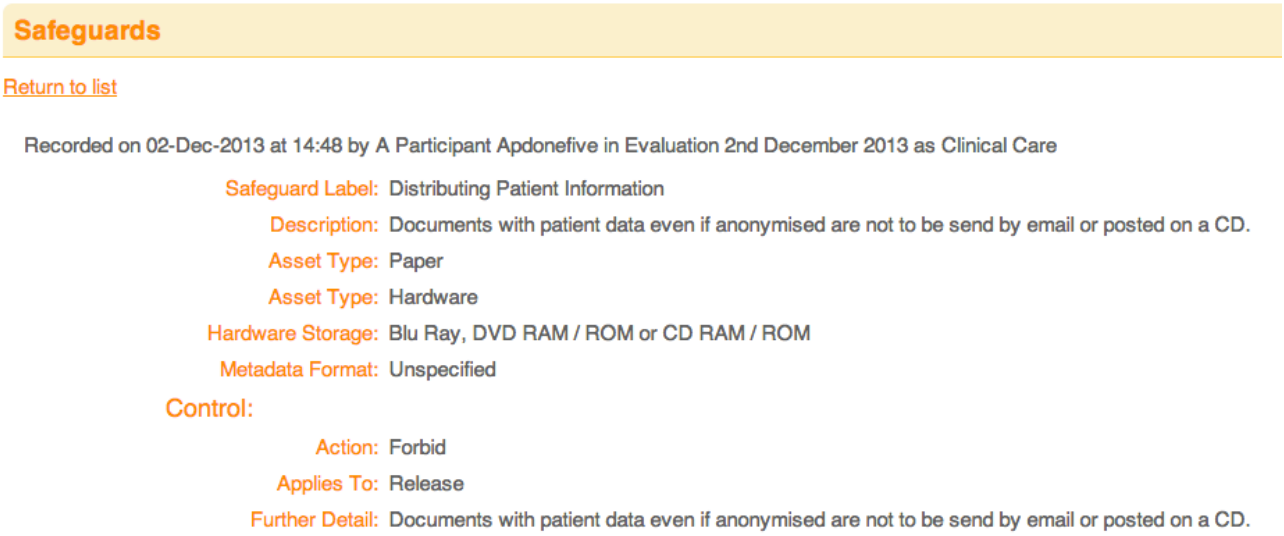
Response Review: the participant was able to use the features of keibi to add details that represented a CD ROM information asset, including a sensible entries for Unique Identifier, Asset Type, Metadata Format and Hardware storage.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.



Response Review: the participant has correctly added the legal basis of consent; the addition of a thorough description shows that the participant has appreciated the meaning of consent a required by the excerpt.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.



Response Review: the participant has correctly added a safeguard that describes the required information. They have labeled it correctly, provided a sensible description, as well as a sensible Asset Type for Hardware. They have also inferred that this Safeguard applies to the Asset Type of paper correctly based on the reference to Document. Hardware Storage and Metadata Format fields are correctly specified. The Control is also correctly specified, including all relevant information in the correct fields. It should be noted that the Participant did not correct the typing error "send" that should have read "sent." The participant did not include the Activity of Sharing Information Assets, and did not include the CD ROM Information Asset for that Activity.

Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.

Response Review: the participant has correctly added the Safeguard in this case, specifying the individual information assets that were available prior to their commencing the experiment. The correct control has also been added. The participant was expected to add an Activity for describing Research when adding this. Additionally, having added the Activity in the next question, they might also have added the Activity to the Safeguard after they had completed the next question.

Question 5: All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.

Safeguards

Return to list

Recorded on 02-Dec-2013 at 15:12 by A Participant Apdonefive in Evaluation 2nd December 2013 as Clinical Care

Safeguard Label: Privacy protection

Description: All research involving human participants, or data samples derived from human participants (such as cohort studies, clinical trials etc.) must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.

Legal Basis Reference: Consent

Activity Reference: CYP IAPT Research

Information Asset Reference: Completed Consent Form

Asset Type: Paper

Asset User Reference: Pinkman, Jesse

Control:

Action: Apply

Applies To: Access

Further Detail: Appropriate consent must be obtained prior to data sharing.

Asset Users

Return to list

Recorded on 02-Dec-2013 at 14:58 by A Participant Apdonefive in Evaluation 2nd December 2013 as Clinical Care

First Name: Jesse

Surname: Pinkman

Job Title: Clinician in Department of Psychological Therapists

Affiliated Organisation: UCL

Responsibility: Responsible for CYP IAPT delivery

Role: Senior Psychoanalyst

Activities

Return to list

Recorded on 02-Dec-2013 at 15:18 by A Participant Apdonefive in Evaluation 2nd December 2013 as Clinical Care

Activity Label: CYP IAPT Research

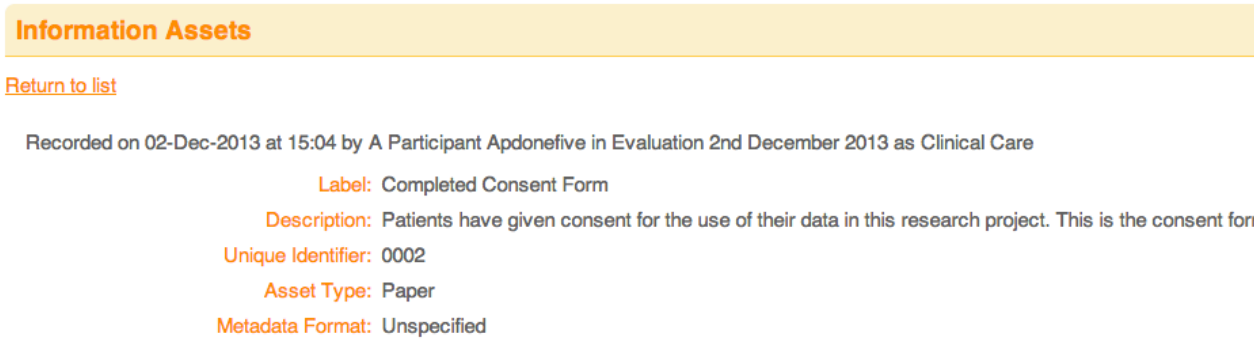
Description: CYP IAPT Research project funded by xPSRC. xPSRC grant number 03203434

Purpose: Medical Research

Information Asset Reference: Completed Consent Form

Information Asset Reference: Mac Mini

Asset User Reference: Pinkman, Jesse



Response Review: the participant has correctly added the Safeguard, with appropriate detail in most fields. They have also added the Activity and Asset User. However, the Asset User and Activity for this question had incorrect details added: specifically naming a research project was incorrect - the Activity should be generic for research (including cohort studies and clinical trials) and not refer to a specific project, particularly one that is not research. Additionally, the Asset User details of Job Title, Responsibility and Role are all inferred and do not relate to a researcher or anything that was asked for in the question or excerpt. The participant has added a Consent Form Information Asset. This has surpassed expectations as only a reference to the Consent Legal Basis provided in Question Two was needed: by providing an Information Asset, this makes the stipulation more specific.

Question 6: The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances.

Response Review: The participant has specified the expected Information Asset and Safeguard correctly for the most part. All of the Mac Mini Details are correct. In the Safeguard however, the specification of the Control is arguably too vague.

The assumption that the Activity this applies to is the CYP IAPT research shows an appreciation of the use of the Mac Mini, but there is nothing in the excerpt to indicate that this stipulation applies to research. There is no PDA Information Asset provided

Information Assets

Return to list

Recorded on 02-Dec-2013 at 15:17 by A Participant Apdonefive in Evaluation 2nd December 2013 as Clinical Care

Label: Mac Mini
Description: Server containing research-related patient data
Unique Identifier: the-mac-address
Asset Type: EHR Server
Metadata Format: openEHR
Hardware Storage: Server

Safeguards

Return to list

Recorded on 02-Dec-2013 at 15:25 by A Participant Apdonefive in Evaluation 2nd December 2013 as Clinical Care

Safeguard Label: Server connectivity
Description: The Mac Mini shall not be networked using the Ethernet connection, Airport wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, computer) shall be connected. The Remote Control function will not be used under any circumstances.
Activity Reference: CYP IAPT Research
Information Asset Reference: Mac Mini
Asset Type: EHR Server
Hardware Storage: Server
Control:
Action: Apply
Applies To: Access
Further Detail: The Mac Mini will remain disconnected from all other sources at all times

Question 7: No copies of data shall be made or distributed from the Mac Mini on any removable medium.

Safeguards

Return to list

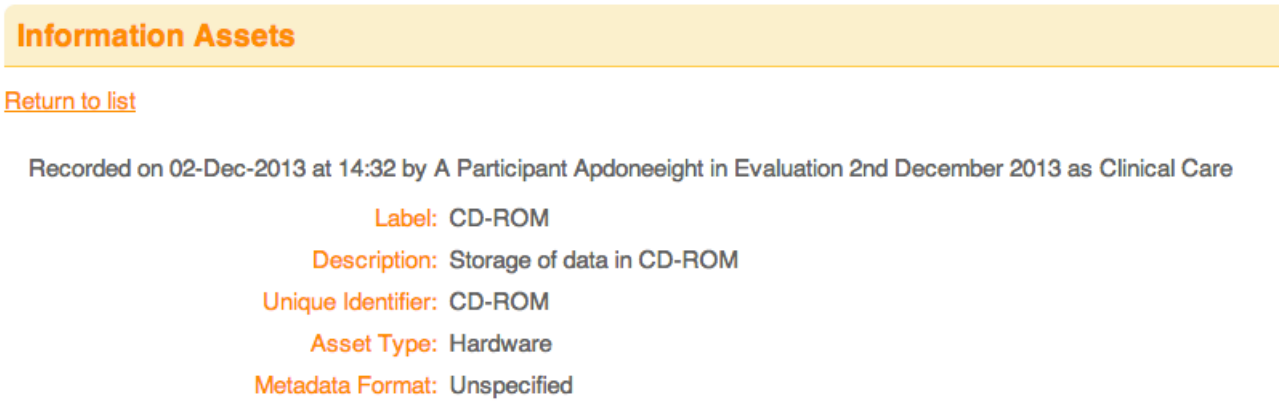
Recorded on 02-Dec-2013 at 15:28 by A Participant Apdonefive in Evaluation 2nd December 2013 as Clinical Care

Safeguard Label: Copying data
Description: No copies of data shall be made or distributed from the Mac Mini on any removable mediu
Activity Reference: CYP IAPT Research
Information Asset Reference: Mac Mini
Asset Type: EHR Server
Hardware Storage: Server
Asset User Reference: Pinkman, Jesse
Control:
Action: Apply
Applies To: Access
Further Detail: No copying permitted

Response Review: The participant has specified the Safeguard correctly, adding the appropriate Information Asset reference. The assumption that the Activity this applies to is the CYP IAPT research shows an appreciation of the Context of Use, in this case a Research Project, though this is unnecessary given the context of use. Also, there is an error here in that the Asset User Jesse Pinkman has been added. According to the policy excerpt, this should apply to any Asset User and not any specific examples of them. The Further Detail for the Control is also too vague. The Information Asset Removable Media was not provided, and the Asset Type of Blu Ray … CD-ROM should have been added as a means of providing reference to removable media.

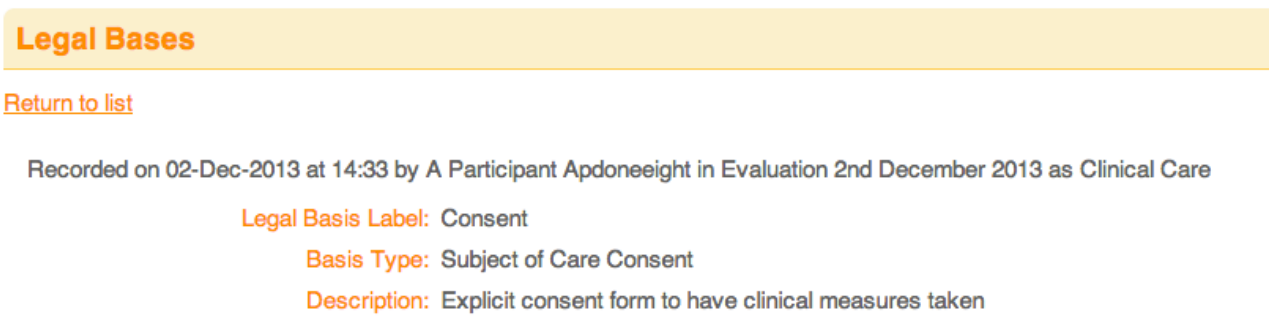
Participant APD_00018

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.



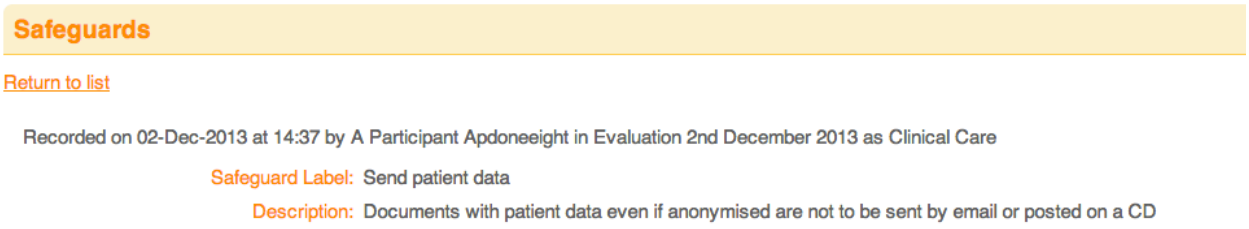
Response Review: The participant has specified the Information Asset details correctly for the most part. An Unique Identifier of CD-ROM however is not uniquely identifying.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.



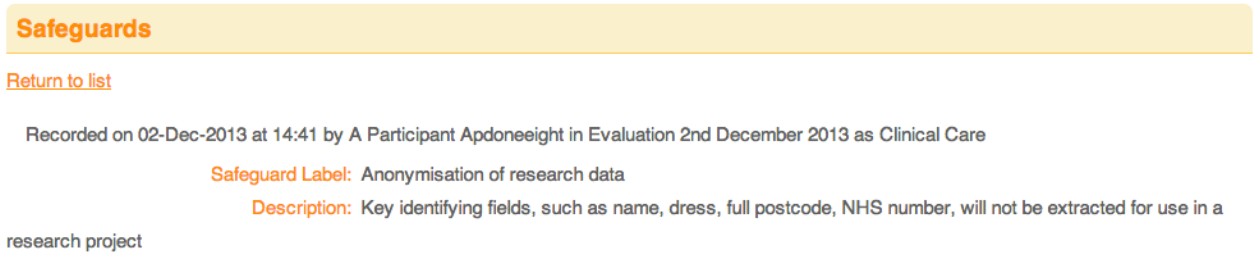
Response Review: The participant has specified the Consent Legal Basis details correctly, though has assumed that this is only for clinical measures to be taken. It is not clear what is meant by clinical measures (whether it refers to a height or weight measurement, or some form of clinical intervention like treatment), and this is not specified in the question.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.



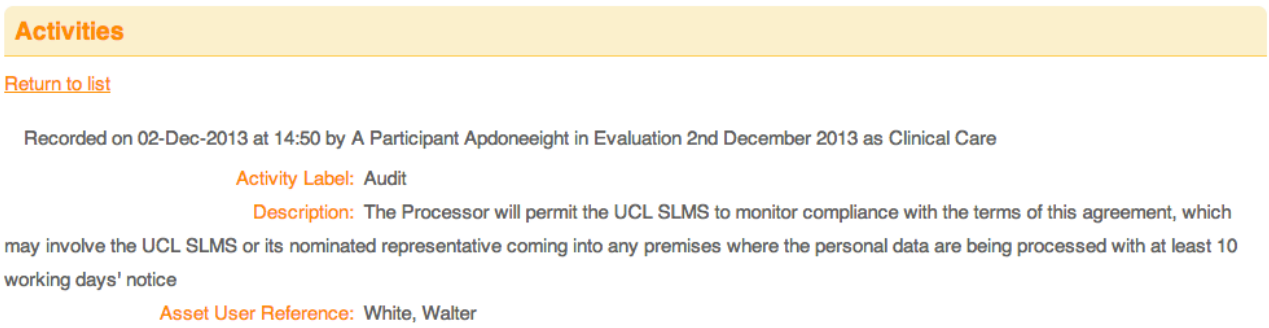
Response Review: The participant has once again specified a Safeguard with a Label and Description. There is no reference to the CD ROM Information Asset, and no Control has been specified. In this case, as with the first question, providing the details of the policy in the description may be sufficient to achieve the correct behaviour from the reader of this policy. Additionally, the Activity of Sharing Information has also not been provided.

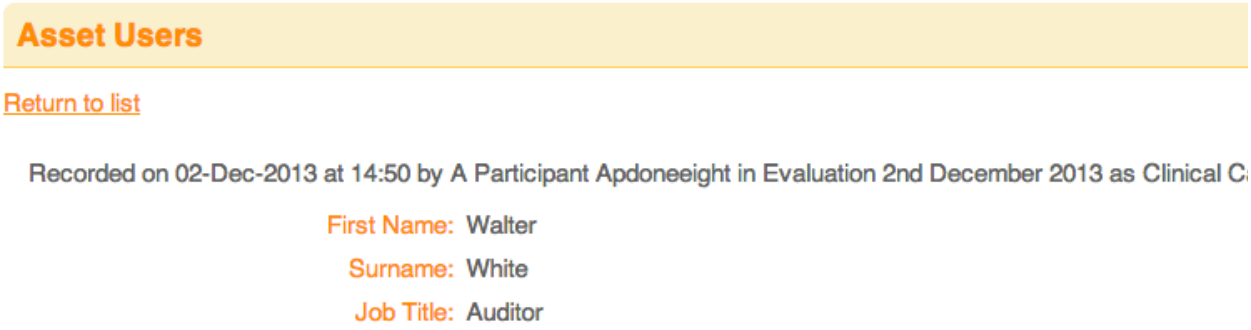
Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project



Response Review: The participant has specified a Safeguard with a Label and Description. Whilst no Control has been specified, this does not necessarily represent a failure to understand what is required. An analysis of how participant APD_00015 answered the question in experiment two would determine if this was sufficient to provide the requisite detail to help them make an effective decision. This will however make computable refinement impossible to achieve. Additionally, the Activity describing Research as not provided.

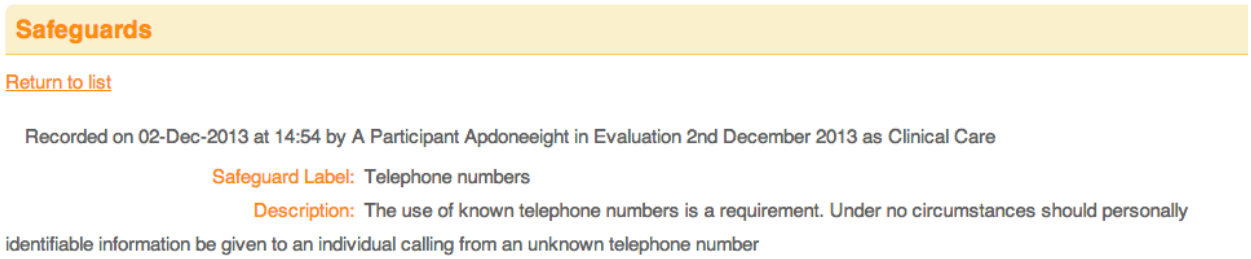
Question 5: Audit: The Processor will permit the UCL SLMS to monitor compliance with the terms of this agreement, which may involve the UCL SLMS or its nominated representative coming onto any premises where the personal data are being processed with at least 10 working days' notice.





Response Review: The participant has specified an Activity and an Asset User as expected with sufficient detail. They have not, however provided a Safeguard, opting instead to place the details of the policy in the Description of the Activity. This approach is not necessarily correct as the stipulations for allowing the Audit Activity should be placed in a Safeguard. It is nevertheless arguable that these stipulations could form part of the detail of the Activity and that would be sufficient for users to access the details and behave appropriately, if they knew to look at the detail of the Audit Activity.

Question 6: The use of known telephone numbers is a requirement. Under no circumstances should personally identifiable information be given to an individual calling from an unknown telephone number.



Response Review: The participant has correctly specified a Safeguard, opting again to add the text of the excerpt into the Description field. Arguably this is not

incorrect, especially given that the excerpt is relatively simple, however it could be argued that they have neglected to enter telephone numbers as an Information Asset and added it to this Safeguard. The Safeguard does not add much detail and this is reflected in the scores below.

Question 7: In the event that unauthorised access was gained through security codes or passes being compromised the Line Manager must immediately take appropriate action to prevent further unauthorised access. Security codes should be changed and any passes used for unauthorised access should be disabled.

Information Assets

Return to list

Recorded on 02-Dec-2013 at 15:00 by A Participant Apdoneeight in Evaluation 2nd December 2013 as Clinical Care

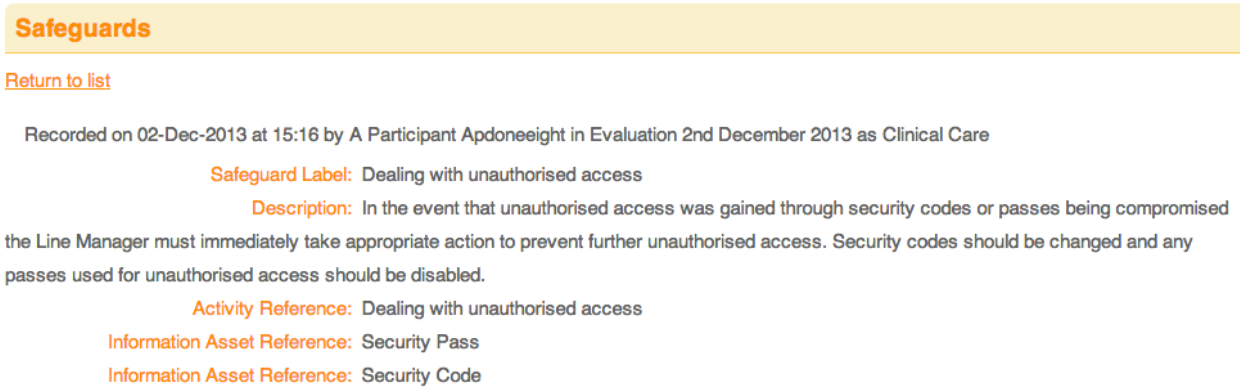
Label: Security Pass
Description: Security Pass
Unique Identifier: Security Pass
Asset Type: Hardware
Metadata Format: Unspecified

Information Assets

Return to list

Recorded on 02-Dec-2013 at 15:00 by A Participant Apdoneeight in Evaluation 2nd December 2013 as Clinical Care

Label: Security Code
Description: Security Code
Unique Identifier: Security Code
Asset Type: Paper
Metadata Format: Unspecified



Response Review: The participant has correctly specified two additional Information Assets, a Security Code and Security Pass. They have assumed here that the Security Code is stored on paper, though this may not be necessarily be correct. Additionally, the Unique Identifier of Security Code and Pass are not particularly unique. The participant specified a Safeguard correctly, but it lacks a specified Control to guide appropriate behaviour as required by the policy excerpt. They have also correctly supplied and referred to an Activity for dealing with unauthorised access. It should be noted that Participant APD_00018 answered the question associated with this excerpt correctly.

Experiment 2

Participant APD_00015

General comments: had to go back to activity and add assets once they are created.

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

I would not agree to the request

I would need more information before allowing any kind of sharing. There is no information in *keibi* about who owns the data (as far as I can see)

I cannot see any explicit Safeguards relating to this item of information.

Response Review: The expected response was that anonymised data could be released, but not shared via email or CD, based upon the stipulations authored by Participant APD_00018 in Questions three and four. Participant APD_00015 did not see the details in the Description fields of the Safeguards prepared by participant APD_00018. Out of the three responses, two are incorrect as a result of participant APD_00015 failing to see the details provided by keibi as authored by APD_00018.

Question 2: You have received a telephone request for names and addresses of three participants in a research trial. How would you respond?

Response to Question 2:

Initial request denied pending further investigation.

Telephone requests must be from a verifiable number. If number seemed valid, I would call back to verify that it is correct.

I would need more information about who may have access to the data. The *keibi* policy would need updating in this respect.

Response Review: This question expected that the participant would use the details added in questions five and six and Participant APD_00015 made two expected responses, and surpassed expectations by suggesting that access control stipulations are added to the policy.

Response to Question 3:

Question 3: A team member from UCL CLMS has asked to come and inspect your compliance with a data sharing agreement you hold with them within the next week. How would you respond to their request?

If I were considered to be acting as the data controller (Principal/Co-Investigator perhaps?) in my organisation I may choose to act without referring to my line manager. Otherwise I would refer.

The *keibi* policy states that UCL SLMS may audit with 10 days notice.

Response Review: Participant APD_00015 made the correct response and provided the correct justification for it. They provided additional details about consulting a line manager using their own expertise since the original excerpt and transposed details in keibi were vague on who would enact the cancellation of codes and passes.

Question 4: You have received a report of an unauthorised access to an information asset. How do you respond to this?

Response to Question 4:

Report to Line Manager.

Undertake any associated actions if requested by Line Manager, regarding changing security codes and passes.

Response Review: Participant APD_00015 made two responses, both of which were correct. The second response inferred the need for Line Manager instruction to disable the codes and the passes, though this is implicit in both the original policy excerpt and the details added on *keibi*.

- *Participant APD_00018*

General Comments:

Difference between Safeguards and Activities?

How to delete an Activity?

How to record transfer of information/data/docs?

Necessary to enter things in different places?

What if all necessary info is in one document?

Why Excerpts and not whole documents?

Substitute of current Excel File used by SLMS for NHS Toolkit?

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

to Question 1:

All the research-related patient data are stored in the local server, which is not networked. Any data contained in this server can only be accessed locally as they

can not be copied to any external device, emailed or posted. Therefore, emailing them in a spreadsheet, even if anonymised, is not an option.

Response Review: This question expected that the participant would use the details added in questions three and four. There are three responses: the first response shows that they have also included the details from questions six and seven, incorrectly assuming that all research related data was being stored on the Mac Mini Server. This is not the case and has not been specified in any of the details added by Participant APD_00015. The second response is correct, though not relevant in this question. The third response is correct, however.

Question 2: You have been asked to transfer data from a Mac Mini holding a research data repository to a colleague using a USB key. How would you respond?

Response to Question 2:

I would reject the request on the same grounds as above.

Response Review: This question expected that the participant would use the details added in questions six and seven by Participant APD_00015. There is a single response, which is correct. There was another response, which could have been made that made clear that there was no indication on how to fulfill the request.

Question 3: A colleague has requested that you network the Mac Mini to perform a critical software update. How would you respond?

Response to Question 3:

Correct response - option 4: I would speak to the policy authors and ask them whether this is an acceptable case for breaching the policy, or how they would proceed.

The request seems reasonable as the software update is very important, but the responsibility for breaching the policy should lie on the people who designed it.

Response Review: This question expected that the participant would use the details added in question six by Participant APD_00015. The response is correct - the participant understood that the software update was important and despite there being no mention of how to achieve this in keibi, was not mislead by the lack of detail and picked the answer that would support best practice.

Question 4: A research colleague has asked for some patient data. They say that it can be anonymised and that they would be able to receive it via email, on a posted CD or they will be happy to collect it on an encrypted USB key. How do you respond?

Response to Question 4:

Patient data are stored on Mac Mini, which is not networked. Copies of data cannot be put on a removable medium, which rules out making a copy onto a USB key or CD. Patient Data cannot be sent by email or posted on a CD either, even if anonymised.

Therefore I would have to respond that this request cannot be fulfilled.

Response Review: This question expected that the participant would use the details added in question four, five, six and seven by Participant APD_00015. There is a single response and it is correct - the participant understood the pertinent details and responded correctly.

11th December 2013

Participant APD_00019

Experiment One

APD_00019

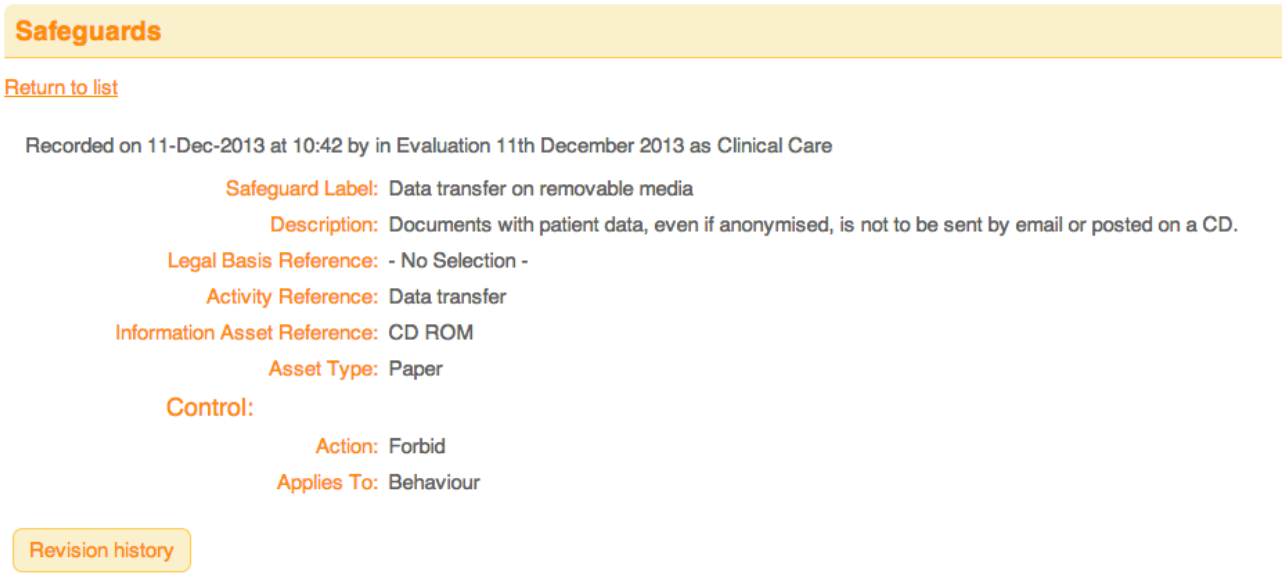
Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.

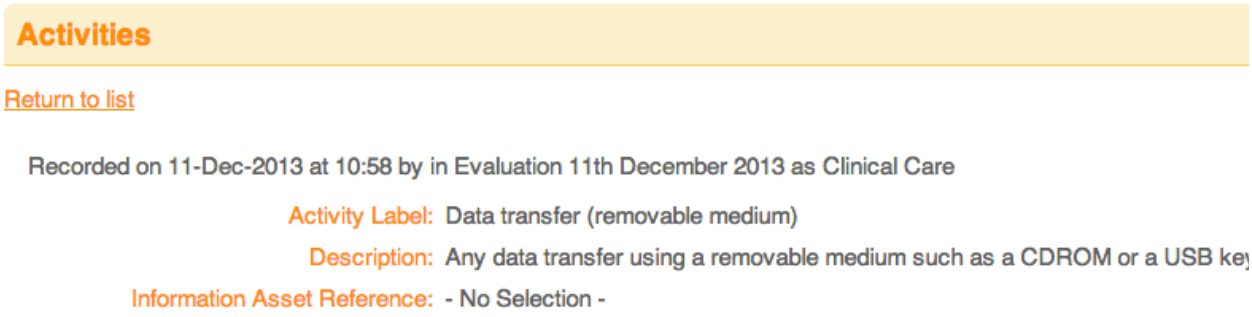
Response Review: the participant was able to use the features of keibi to add details that represented a CD ROM Information Asset.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.

Response Review: the participant has correctly added the legal basis of Consent.

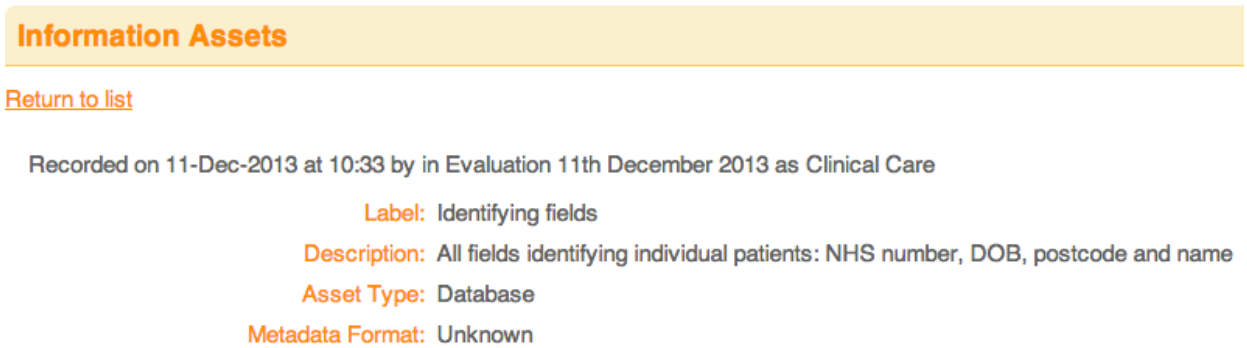
Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.

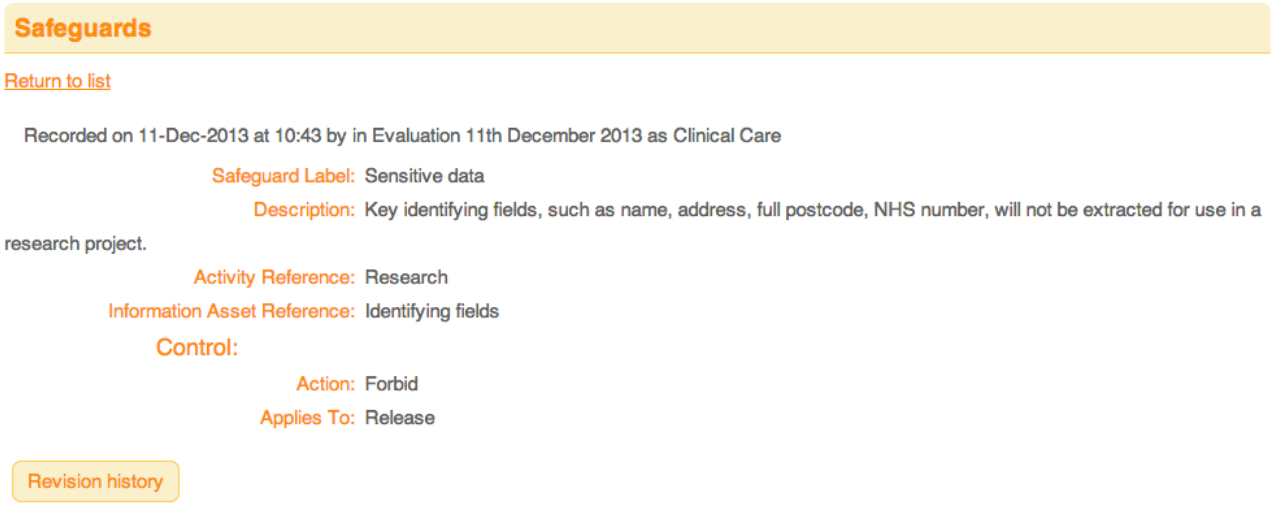


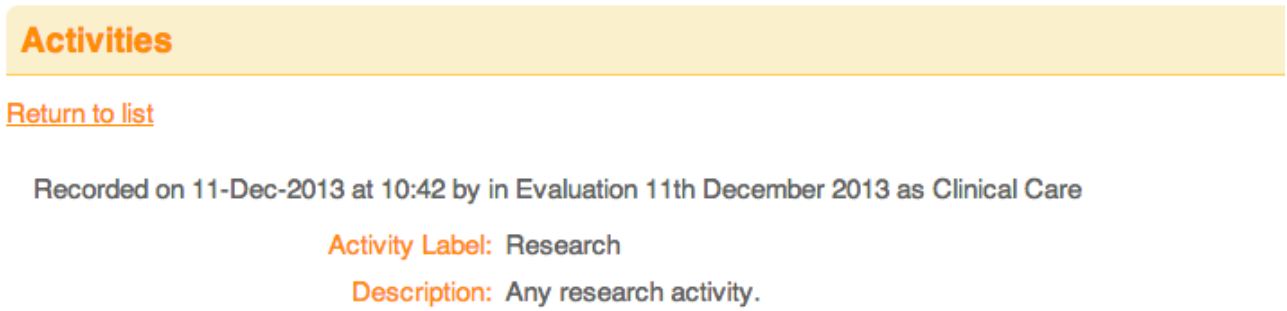


Response Review: the participant has correctly added a Safeguard that describes the required information. They have labeled it correctly, provided a reasonable description (some of the wording of the original excerpt). They have inferred that this Safeguard applies to the Asset Type of paper correctly based on the reference to Document, but have not applied the Asset Type of Hardware to account for the CD ROM. They correctly refer to the CD ROM Information Asset and Data Transfer Activity (as noted below). There is an extraneous reference to a Legal Basis. The Control is also correctly specified, including relevant information in the correct fields. The participant also included the Activity for Transferring Data, with Sensible details (though this includes an Information Asset Reference of No Selection). They have also not provided an email Activity.

Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.



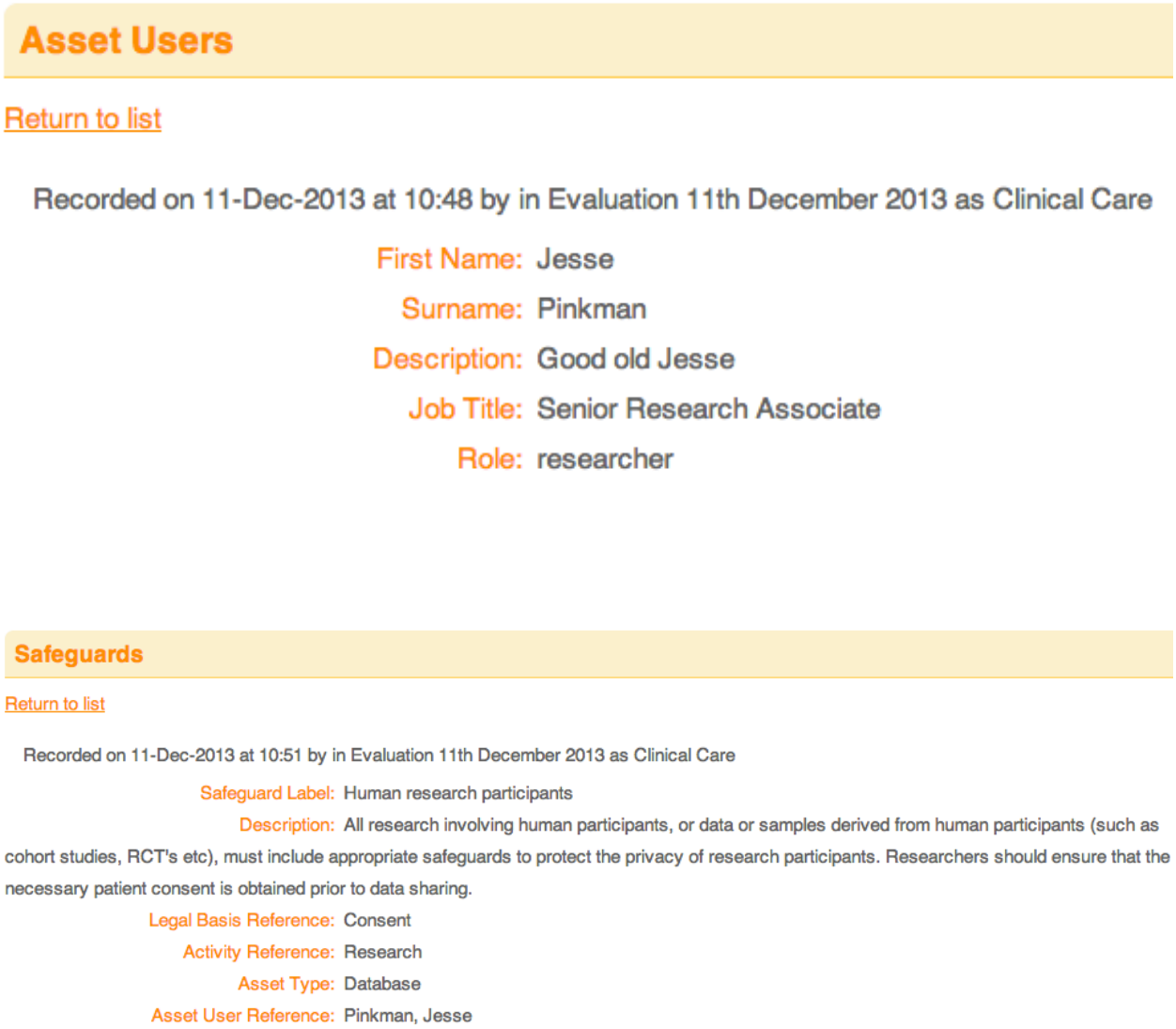




Response Review: the participant has correctly added the Safeguard in this case for the most part, but has opted to create an Information Asset for all the identifying fields unnecessarily, ignoring the available Information Assets that had been pre-prepared. They have nevertheless referred to this Information Asset in the Safeguard, and have referred to an Activity for Research, which they have also prepared (see below). The correct Control has also been added. The participant was expected to add an Activity for describing Research and has done this correctly.

Question 5: All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include

appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.



Asset Users

Return to list

Recorded on 11-Dec-2013 at 10:48 by in Evaluation 11th December 2013 as Clinical Care

First Name: Jesse
Surname: Pinkman
Description: Good old Jesse
Job Title: Senior Research Associate
Role: researcher

Safeguards

Return to list

Recorded on 11-Dec-2013 at 10:51 by in Evaluation 11th December 2013 as Clinical Care

Safeguard Label: Human research participants
Description: All research involving human participants, or data or samples derived from human participants (such as cohort studies, RCT's etc), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.
Legal Basis Reference: Consent
Activity Reference: Research
Asset Type: Database
Asset User Reference: Pinkman, Jesse

Response Review: the participant has correctly added the Safeguard, with appropriate detail in most fields. They have correctly added a reference to the Research Activity, the Consent Legal Basis as well as the Asset User that they have authored in this question (see below). They have added a reference to the Asset Type of database, but this is not indicated in the excerpt. They have also not added a Control, relying instead on the details in the description field as participant

APD_00018 did. They have added an Asset User with correct details, though the Description field is not that descriptive!

Question 6: The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances.

Safeguards

Return to list

Recorded on 11-Dec-2013 at 11:02 by in Evaluation 11th December 2013 as Clinical Care

Safeguard Label: Networking
Description: The Mac Mini shall not be networked using Ethernet connection, Airport, Wireless card, Firewire or USB - it shall remain in a non-networked state. No other device (PDA or other computer) shall be connected. The Remote Control function will not be used under any circumstances.
Activity Reference: Computer networking
Information Asset Reference: All computing devices
Asset Type: Hardware
Control:
Action: Forbid
Applies To: Behaviour
Further Detail: Keep devices off the network.

Revision history

Information Assets

Return to list

Recorded on 11-Dec-2013 at 10:52 by in Evaluation 11th December 2013 as Clinical Care

Label: All computing devices
Description: Any mobile or non mobile computing devices such as laptops, PDA, computers.
Asset Type: Hardware

Information Assets

Return to list

Recorded on 11-Dec-2013 at 10:56 by in Evaluation 11th December 2013 as Clinical Care

Label: Mac Mini

Description: Jesse's Apple Mac Mini

Asset Type: Hardware

Activities

Return to list

Recorded on 11-Dec-2013 at 10:58 by in Evaluation 11th December 2013 as Clinical Care

Activity Label: Data transfer (removable medium)

Description: Any data transfer using a removable medium such as a CDROM or a USB ke

Information Asset Reference: - No Selection -

Activities

Return to list

Recorded on 11-Dec-2013 at 10:53 by in Evaluation 11th December 2013 as Clinical Care

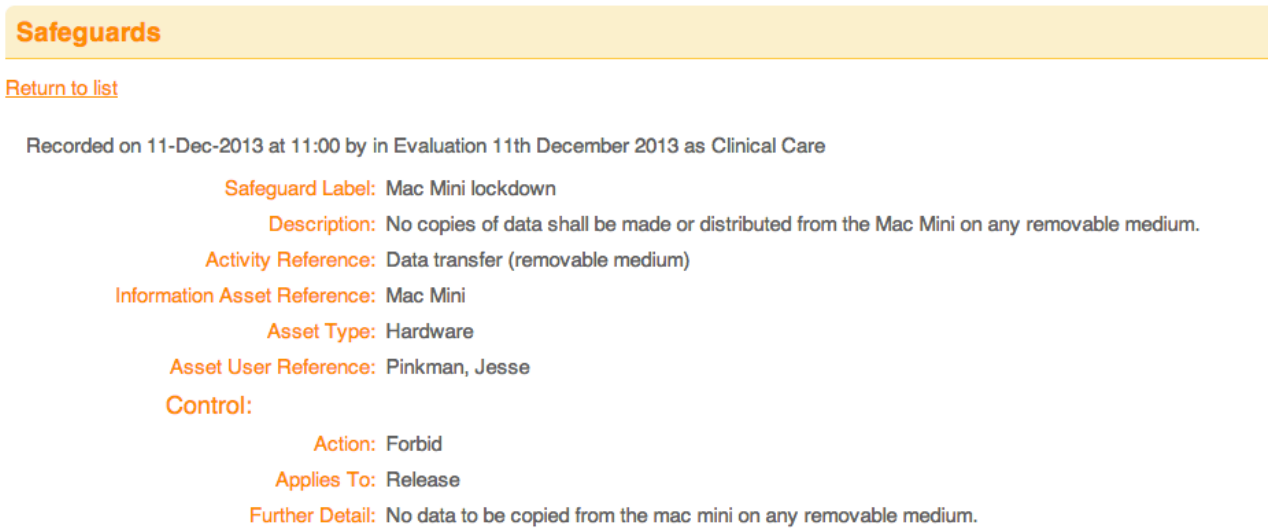
Activity Label: Computer networking

Description: Any transfer of data to or from a device using a computer network.

Response Review: The participant has specified the expected Information Assets and Safeguard correctly for the most part. They have provided a Networking Safeguard that was well specified with correct details, but has included a reference to an Information Asset for All Computing Devices: it stipulates that all devices must be kept off the network, which is incorrect: only the Mac Mini should remain off the Network. The reference to the Networking Activity is also correct. The

response assumes that this refers to the Asset User Jesse Pinkman only, however. This suggests that the participant has in this case inferred that scenario that involves Jesse Pinkman dictates the basis for the policy authoring, though this is not necessarily correct. All of the Mac Mini Information details are correct, with the exception that it belongs to the Asset User Jesse Pinkman. There is also an Information Asset added for all computing devices, which is correct, and includes sufficient details. The participant has also added a more specific Activity of Transferring Data on a removable medium, with reasonable details than the transferring data Activity specified earlier, and an unexpected Activity for Networking, with reasonable details.

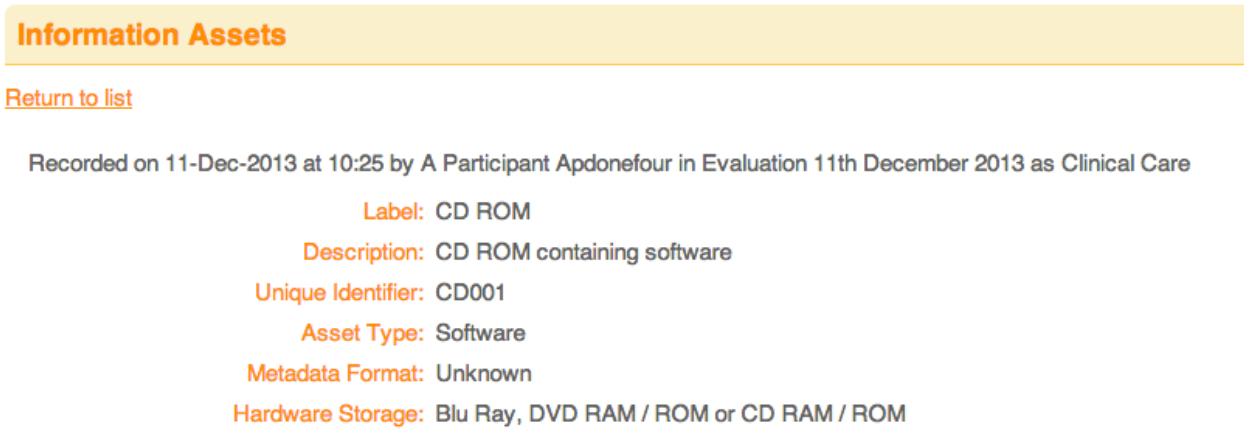
Question 7: No copies of data shall be made or distributed from the Mac Mini on any removable medium.



Response Review: The Safeguard for locking down the Mac Mini is also correct, with the correct reference to the Mac Mini Information Asset. There is an incorrect reference to Jesse Pinkman, picking up the error that was made when the Mac Mini Information Asset was specified in the previous question. There is also the omission of the CD ROM Information Asset reference.

Participant APD_00014

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.



Response Review: The participant has specified the Information Asset details.

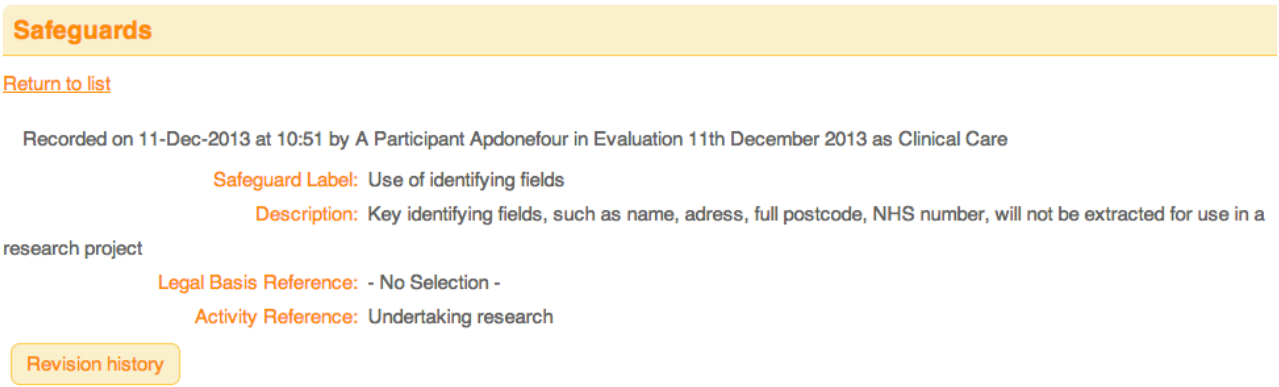
Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.

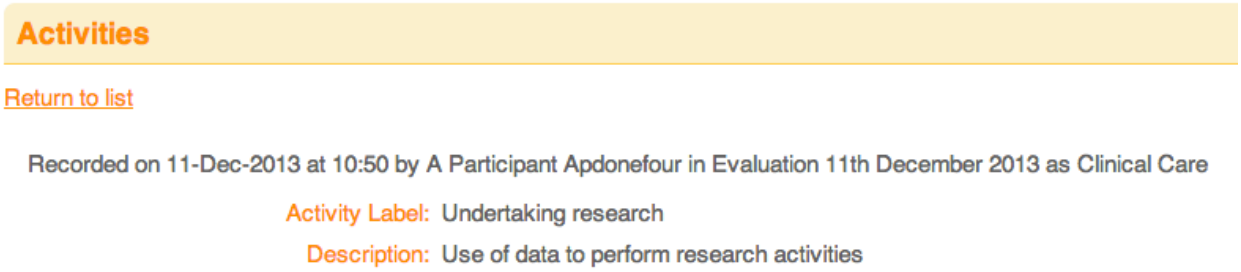
Response Review: The participant did not respond to this question.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.

Response Review: The participant has specified a Safeguard with correct details, though there is no reference to the CD ROM Information Asset . Additionally, the Activity of Sharing Information has also not been provided.

Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.





Response Review: The participant has specified a Safeguard with a Label and Description, but no Control and no reference to the identifying fields that were added. An analysis of how participant APD_00019 answered the question in experiment two would determine if this was sufficient to provide the requisite detail to help them make an effective decision. This will however make computable refinement impossible to achieve. Additionally, there is a No Selection Reference for the Legal Basis, but has included a correct reference to the Activity.

Question 5: Audit: The Processor will permit the UCL SLMS to monitor compliance with the terms of this agreement, which may involve the UCL SLMS or its nominated representative coming onto any premises where the personal data are being processed with at least 10 working days' notice.

Asset Users

Return to list

Recorded on 11-Dec-2013 at 10:57 by A Participant Apdonefour in Evaluation 11th December 2013 as Clinical Care

First Name:	Walter
Surname:	White
Description:	Research staff
Job Title:	Chemist
Affiliated Organisation:	UCLH
Responsibility:	Information Asset Owner
Role:	Principal Investigator

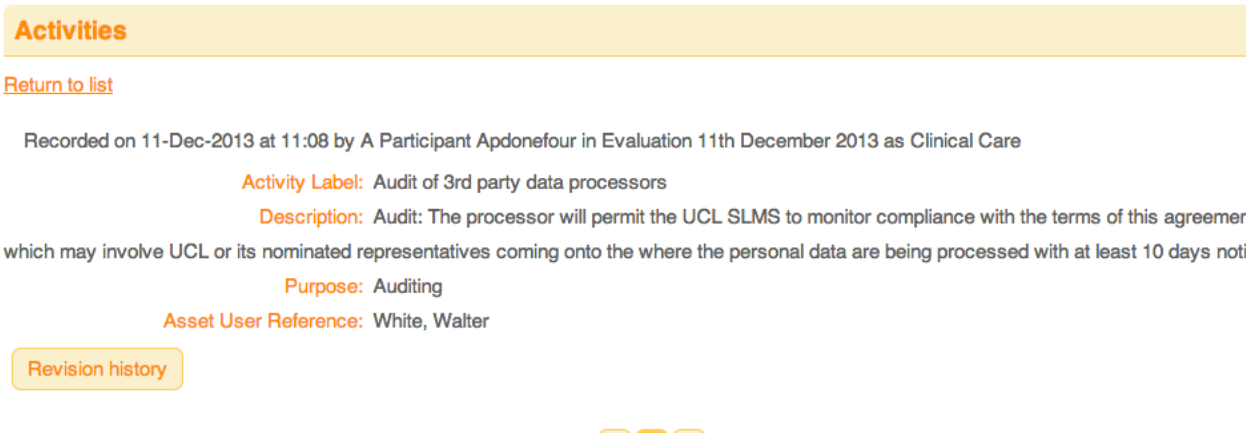
Revision history

Legal Bases

Return to list

Recorded on 11-Dec-2013 at 10:56 by A Participant Apdonefour in Evaluation 11th December 2013 as Clinical Care

Legal Basis Label:	Confidentiality clause
Basis Type:	Employment Contract
Description:	Confidentiality clause in employment contract
Evidence URI:	http://XYZ

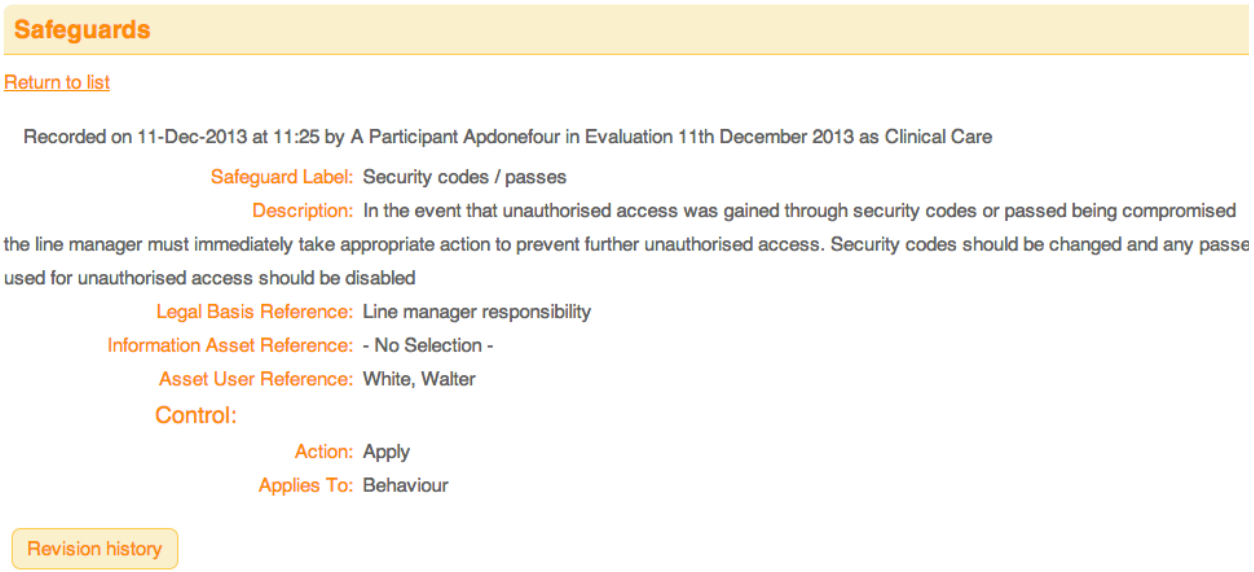


Response Review: As with participant APD_00018, this participant has specified an Activity and an Asset User as expected with sufficient detail. They have not, however provided a Safeguard, opting instead to place the details of the policy in the Description of the Activity. This approach is not necessarily correct as the stipulations for allowing the Audit Activity should be placed in a Safeguard. It is nevertheless arguable that these stipulations could form part of the detail of the Activity and that would be sufficient for users to access the details and behave appropriately, if they knew to look at the detail of the Audit Activity. The Activity details are correct, however the Asset User details have had a number of details inferred: The inference that Walter White is an Information Asset Owner is reasonable, and that their job title is Chemist given the fictional character upon which the Asset User was based. Additionally, it is not unreasonable to assume that they are a Principal Investigator or research staff, and whilst the participant has brought their own expertise to authoring this Asset User, there is no reference to their being an Auditor, which is an omission. The participant has also added a Legal Basis, describing a confidentiality clause. This was unexpected, but shows that the participant was applying their own exerts to answering this question.

Question 6: The use of known telephone numbers is a requirement. Under no circumstances should personally identifiable information be given to an individual calling from an unknown telephone number.

Response Review: The participant has correctly specified a Safeguard, with reasonable details. They have also provided an unexpected Activity for contact by telephone, though this is not just for the purposes of Clinical Trials. There is also no reference to an Information Asset, and they have not supplied an Information Asset for known telephone numbers..

Question 7: In the event that unauthorised access was gained through security codes or passes being compromised the Line Manager must immediately take appropriate action to prevent further unauthorised access. Security codes should be changed and any passes used for unauthorised access should be disabled.



Response Review: The participant has correctly specified a Safeguard with reasonable details, though a No Selection for the Information Asset reference: they did not add the Information Assets for Security Codes and Passes. They have however correctly and unexpectedly added a Line Manager Responsibility Legal Basis, and correctly referred to it in the Safeguard. It should be noted that

Participant APD_00018 answered the question associated with this excerpt correctly.

Experiment 2

Participant APD_00019

General Comments:

"consumable" as asset type

Unclear on Unique Identifiers

Safeguards |:| with assets?

You already know the assets type, why as again?

"Supplied Identifier is not valid for given property" - mark incorrect fields.

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

Despite the fact that the data are anonymised, data transfer over email is not allowed by the policy.

In reality, I would encrypt the data, send the key over SMS and put the encrypted file on UCL Dropbox or IDHS.

Response Review: The expected response was that anonymised data could be released, but not shared via email or CD, based upon the stipulations authored by Participant APD_00014 in Questions three and four. The question was answered correctly with two responses, both of which were correct and provided a reasonable alternative.

Question 2: You have received a telephone request for names and addresses of three participants in a research trial. How would you respond?

Response to Question 2:

Provided that the number in the caller ID is known, I would disclose the information as it's allowed by the policy.

In reality, I would not disclose any PID unless I knew the caller personally. If that's not feasible, I would ask to call their line manager.

Response Review: This question expected that the participant would use the details added in questions five and six and Participant APD_00014 made two responses, both of which were correct and expected.

Question 3: A team member from UCL CLMS has asked to come and inspect your compliance with a data sharing agreement you hold with them within the next week. How would you respond to their request?

Response to Question 3:

The participant chose option 4: *I would agree and offer dates ten or more days later.* This was the correct response.

Policy says "at least" ten days notice.

In reality, if I can do next week I would do so unless I needed the time to go over things internally.

SOURCE: Activity - audit.

Response Review: Participant APD_00014 made the correct response and provided the correct justification for it. They specified that they used the Activity details to answer the question correctly, which, in this case, negated the need for the Safeguard that participant APD_00019 did not provide.

Question 4: You have received a report of an unauthorised access to an information asset. How do you respond to this?

Response to Question 4:

a) Notify Line Manager
b) change security codes / passwords
c) disable accounts

In Reality:

d) be scared
e) notify Line Manager
f) disconnect PC from network / stop service
g) inform UCL Security / ISD Security
h) inform data providers

Response Review: Participant APD_00015 made three responses, all of which were correct. They also exceeded expectations and used their own expertise with four additional responses.

Participant APD_00014
General Comments:

Basis the says NIGB Exemption (now HSCIC, so better to use Section 251)

On Activities - error message 'supplied identifier is not valid for the given property' but no indication of what it refers to.

Information Asset reference - too granular

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

Activities - Data Transfer (hoped to see some info about what is permissible, but dead end).

Safeguards
Could not find advice relevant to this query.
would remove identifiers and ensure remaining data could not be traced back to an individual before transfer (using AES256 and password transferred by separate mechanism).

Response Review: This question expected that the participant would use the details added in questions three and four. There are three responses: the first response stated that they could not see what was permissible in the Activity, where they were looking in the wrong place. The second response said that they

couldn't find relevant details. The final response was for the participant to apply their own expertise answering the question. The participant has not been able to infer a response from the details authored by participant APD_00019. This suggests that the excerpt was too vague.

Question 2: You have been asked to transfer data from a Mac Mini holding a research data repository to a colleague using a USB key. How would you respond?

Response to Question 2:

Safeguard Mac Mini Lockdown

forbids data to be copied on removable medium.

Response Review: This question expected that the participant would use the details added in questions six and seven by Participant APD_00018 and refers to the expected Safeguard. There is a single response, which is correct. There was another response, which could have been made that made clear that there was no indication on how to fulfill the request.

Question 3: A colleague has requested that you network the Mac Mini to perform a critical software update. How would you respond?

Response to Question 3:

Incorrect answer - citing safeguard 'Networking'.

Response Review: This question expected that the participant would use the details added in question six by Participant APD_00015. The response is incorrect - the participant abided by the stipulation in keibi but did not feel that they could question the stipulations that had been stored there. This suggests a reliance on the tool's authority.

Question 4: A research colleague has asked for some patient data. They say that it can be anonymised and that they would be able to receive it via email, on a posted CD or they will be happy to collect it on an encrypted USB key. How do you respond?

Response to Question 4:

Assuming data is held on Mac Mini:

Safeguard 'Data Transfer on removable media' forbids patient data, even if anonymised, from transfer by CD.

Safeguard 'Mac Mini lockdown'

Forbids data transfer by removable medium (e.g. USB Key).

Safeguard 'Networking'

Forbids Mac Mini to be networked therefore, email cannot be used (or data transferred or to be emailed from another machine).

Response: No.

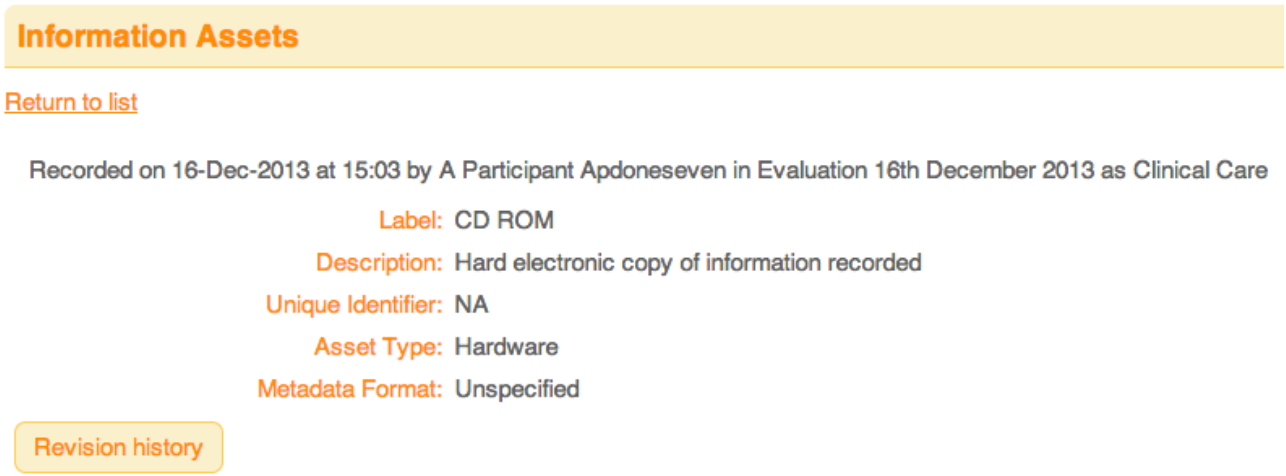
Response Review: This question expected that the participant would use the details added in question four, five, six and seven by Participant APD_00015. There is a single response and it is correct - the participant understood the pertinent details and responded correctly, referring to the appropriate Safeguards.

Evaluation Session 16th December 2013

Experiment One

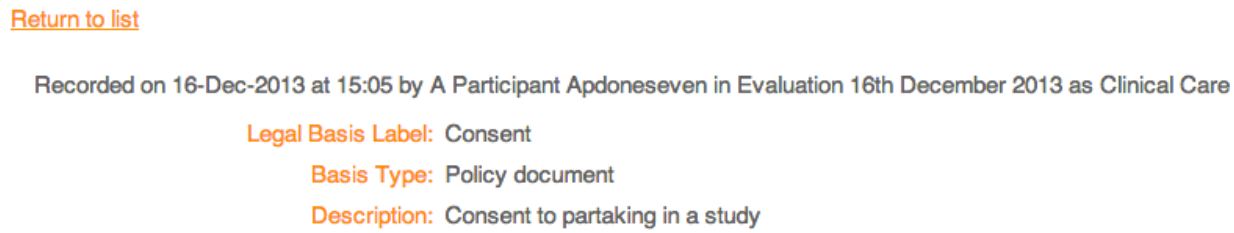
Participant APD_00017

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.



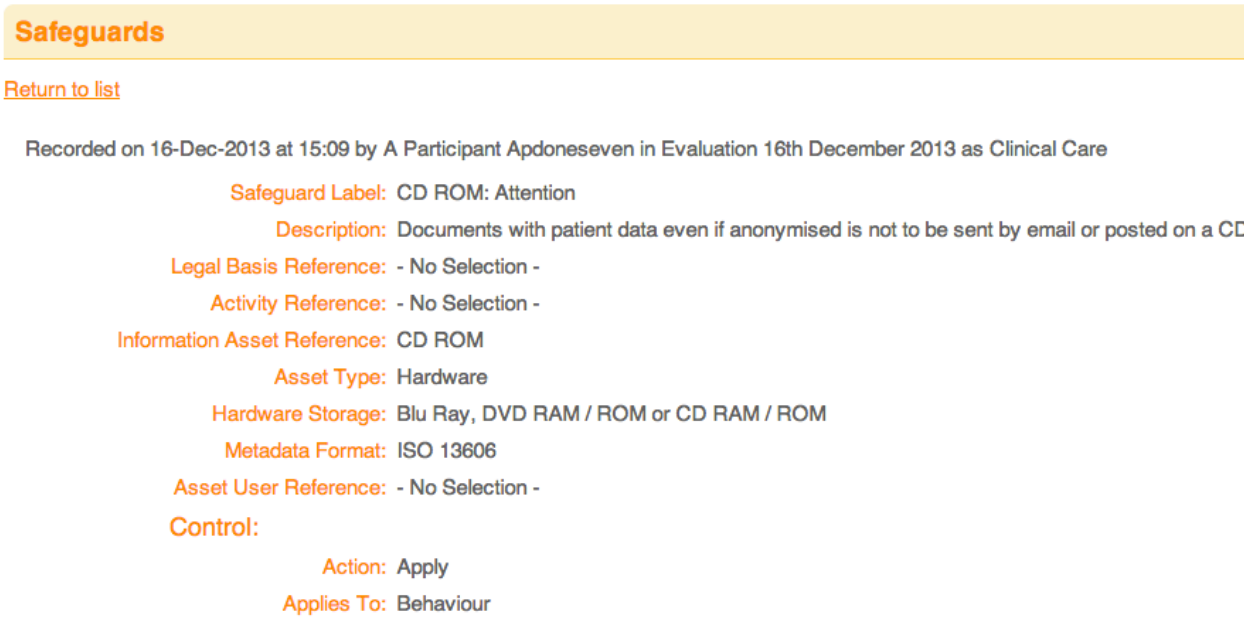
Response Review: the participant was able to use the features of keibi to add details that represented a CD ROM Information Asset.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.



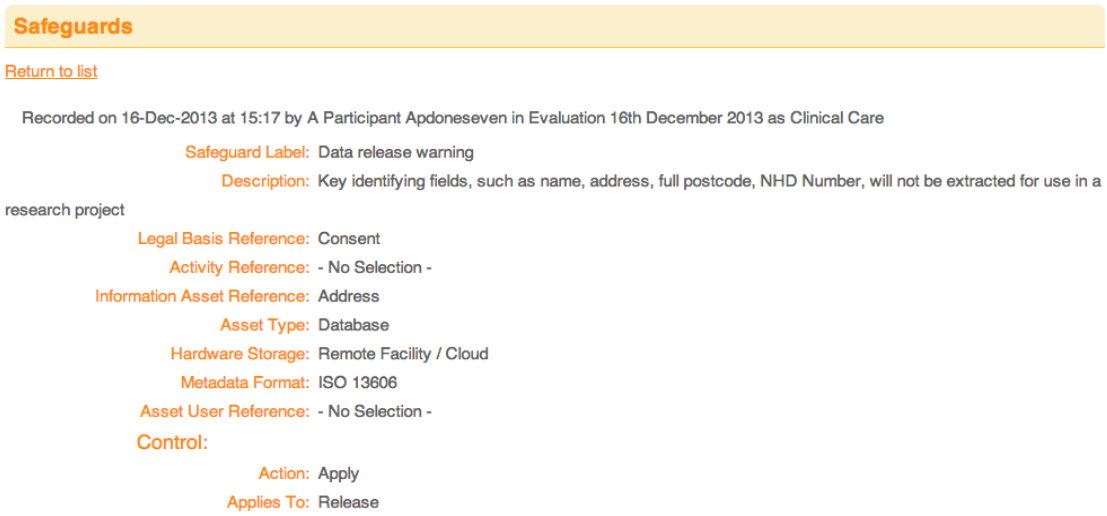
Response Review: the participant has correctly added the legal basis of Consent.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.



Response Review: the participant has correctly added a Safeguard that describes the required information. They have labeled it correctly, provided a reasonable description (some of the wording of the original excerpt). They have inferred that this Safeguard applies to the Asset Type of hardware correctly based on the reference to Document, but have not applied the Asset Type of paper to account for the reference to Document. They correctly refer to the CD ROM Information Asset but have applied Data Transfer Activity (as noted below). There is an extraneous reference to a Legal Basis. The Control is also correctly specified, including relevant information in the correct fields. The participant has not provided the Activities for Research or email.

Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.



Safeguards

Return to list

Recorded on 16-Dec-2013 at 15:17 by A Participant Apdoneseven in Evaluation 16th December 2013 as Clinical Care

Safeguard Label: Data release warning

Description: Key identifying fields, such as name, address, full postcode, NHD Number, will not be extracted for use in a research project

Legal Basis Reference: Consent

Activity Reference: - No Selection -

Information Asset Reference: Address

Asset Type: Database

Hardware Storage: Remote Facility / Cloud

Metadata Format: ISO 13606

Asset User Reference: - No Selection -

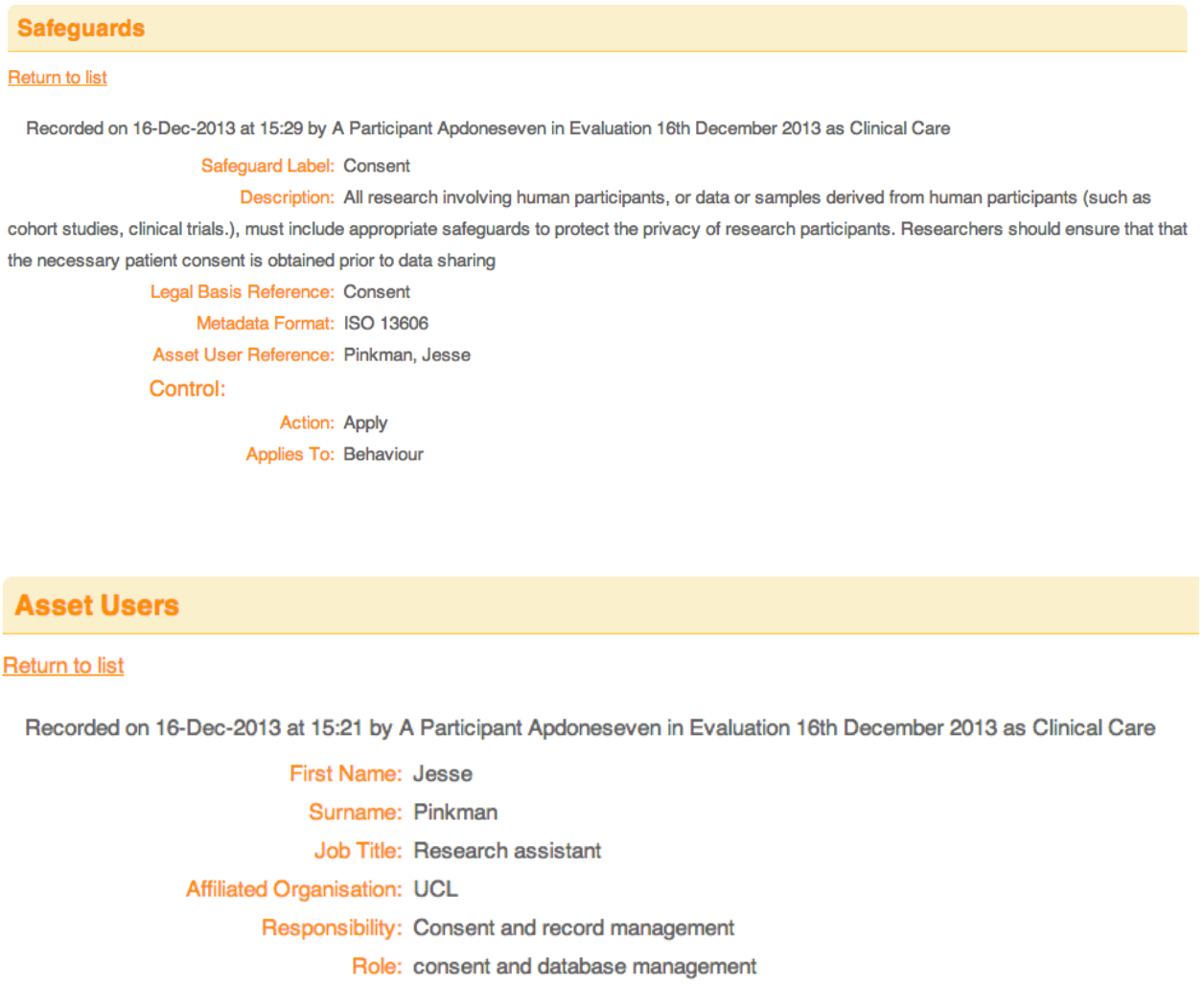
Control:

Action: Apply

Applies To: Release

Response Review: the participant has correctly added the Safeguard in this case for the most part, but has only added the Address Information Asset that had been pre-prepared. the available Information Assets that had been pre-prepared. They have also not referred to a Research Activity or added it. They have also referred to ISO 13606 as a metadata standard, which is not indicated in the excerpt.

Question 5: All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.



Safeguards

Return to list

Recorded on 16-Dec-2013 at 15:29 by A Participant Apdoneseven in Evaluation 16th December 2013 as Clinical Care

Safeguard Label: Consent
Description: All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials.), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that that the necessary patient consent is obtained prior to data sharing
Legal Basis Reference: Consent
Metadata Format: ISO 13606
Asset User Reference: Pinkman, Jesse
Control:
Action: Apply
Applies To: Behaviour

Asset Users

Return to list

Recorded on 16-Dec-2013 at 15:21 by A Participant Apdoneseven in Evaluation 16th December 2013 as Clinical Care

First Name: Jesse
Surname: Pinkman
Job Title: Research assistant
Affiliated Organisation: UCL
Responsibility: Consent and record management
Role: consent and database management

Response Review: the participant has correctly added the Safeguard, with appropriate detail in most fields. They have correctly added a reference to the Consent Legal Basis as well as the Asset User that they have authored in this question (see below). They have added a reference to the Metadata Format of ISO 13606, but this is not indicated in the excerpt. They have added an Asset User with reasonable details.

Question 6: The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances.

Information Assets

Return to list

Recorded on 16-Dec-2013 at 15:35 by A Participant Apdoneseven in Evaluation 16th December 2013 as Clinical Care

Label: Mac Mini

Description: The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB- it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances

Asset Type: Hardware

Metadata Format: ISO 13606

Hardware Storage: Server

Response Review: The participant has specified the expected Information Asset correctly. It is reasonable to assume that the Mac Mini would store an EHR server according to an ISO 13606 format. They have not, however, specified the PDA Information Asset. They have also failed to supply the non networking Safeguard.

Question 7: No copies of data shall be made or distributed from the Mac Mini on any removable medium.

Safeguards

Return to list

Recorded on 16-Dec-2013 at 15:39 by A Participant Apdoneseven in Evaluation 16th December 2013 as Clinical Care

Safeguard Label: Mac Mini

Description: No copies of data shall be made or distributed from the Mac Mini on any removable medium

Information Asset Reference: Mac Mini

Asset Type: EHR Server

Hardware Storage: Server

Control:

Action: Apply

Applies To: Behaviour

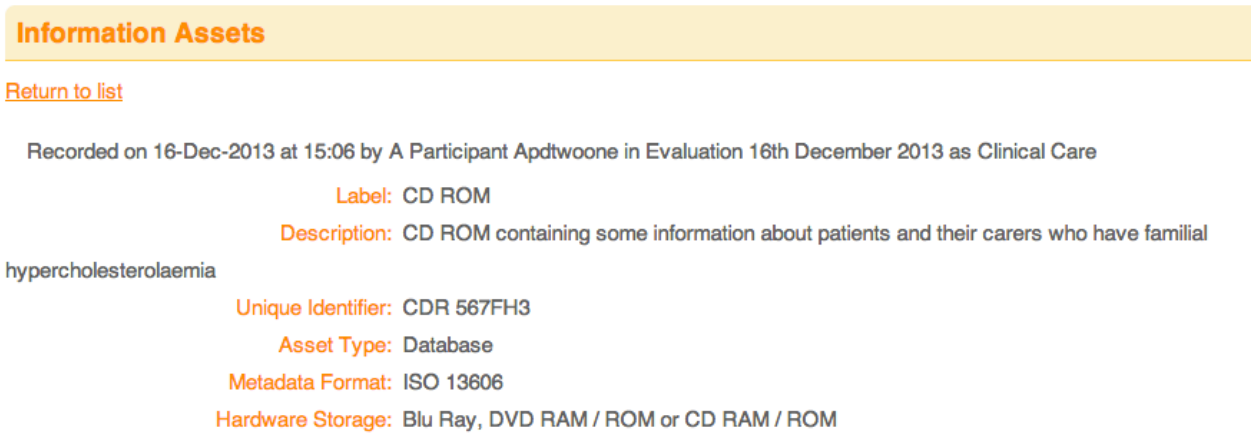
Response Review: The Safeguard for sharing data from the Mac Mini is correct, with the correct reference to the Mac Mini Information Asset. There is also the omission of the CD ROM Information Asset reference.

Participant APD_00021

The responses from participant APD_00021 are provided below.

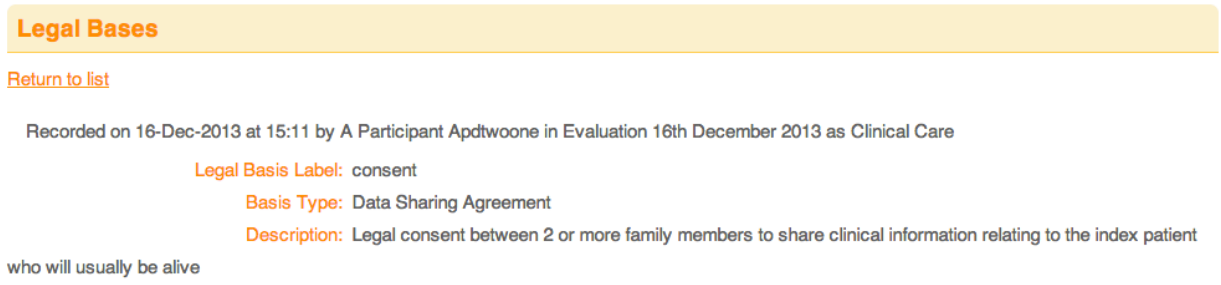
Experiment One

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.



Response Review: the participant was able to use the features of *keibi* to add details that represented a CD ROM Information Asset. The details about what is stored on it are extraneous.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.



Response Review: the participant has correctly added the legal basis of Consent.

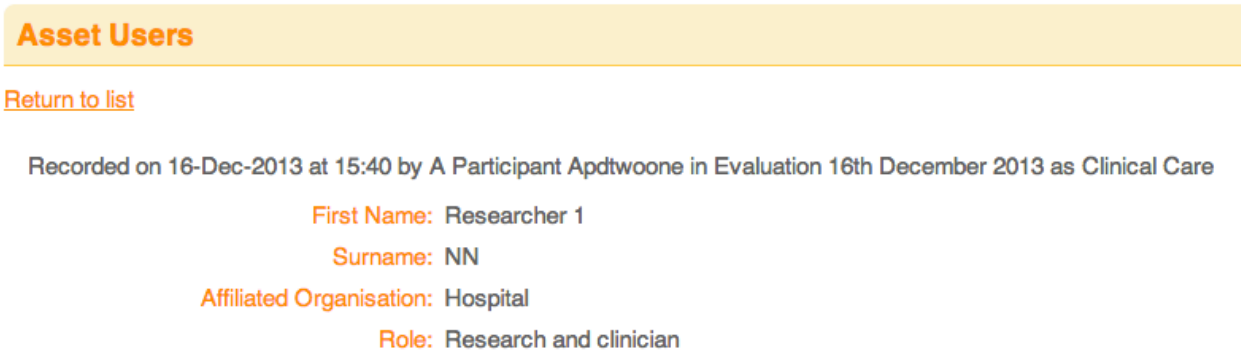
Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.

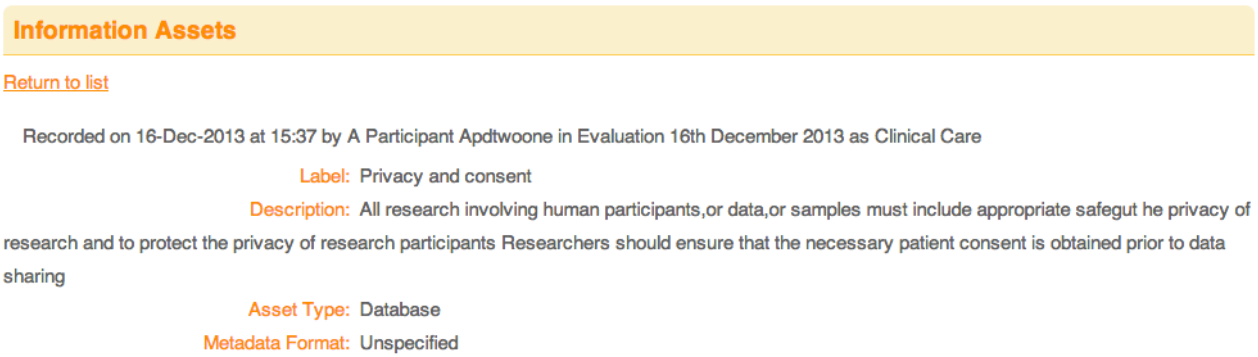
Response Review: the participant has only supplied a Legal Basis, but no Safeguard, or the Activities for Research or email. This was due to their having difficulties with saving the Safeguard, as described in the Group Discussion results below. The participant navigated away from the Safeguard editing screen and before saving it. Since they had not been warned about this issue prior to starting to use the tool, this failure has been put down to a tool error as opposed to participant error.

Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.

Response Review: the participant has correctly provided part of the Safeguard as expected. They have not specified a Control, however, and incorrectly specified Paper as an Asset Type. They have not added the Research Activity either.

Question 5: All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.





Information Assets

Return to list

Recorded on 16-Dec-2013 at 15:37 by A Participant Apdtwoone in Evaluation 16th December 2013 as Clinical Care

Label: Privacy and consent

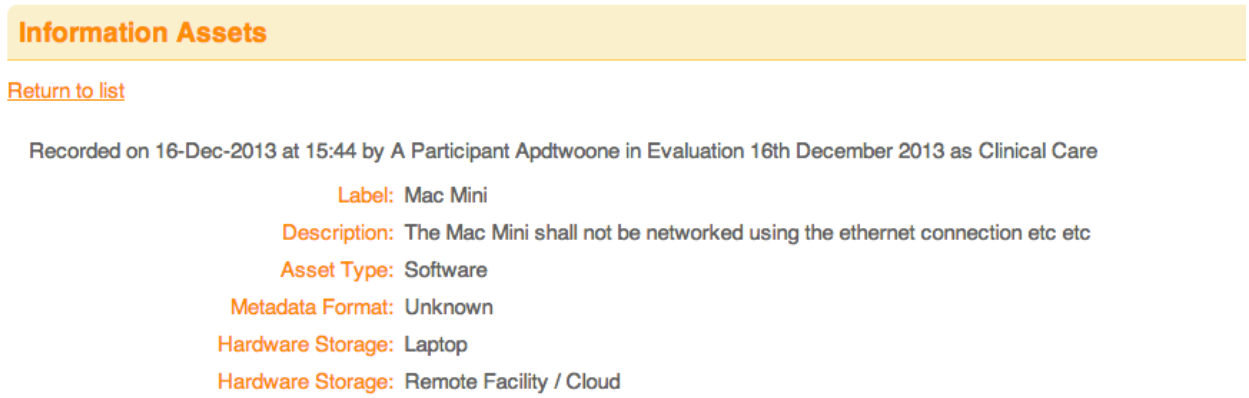
Description: All research involving human participants,or data,or samples must include appropriate safegut he privacy of research and to protect the privacy of research participants Researchers should ensure that the necessary patient consent is obtained prior to data sharing

Asset Type: Database

Metadata Format: Unspecified

Response Review: the participant has added an Asset User as expected, but has not provided correct details as directed by the question. They have sought to de-identify the researcher by applying their knowledge of the area. This defeats the purpose of having named Asset Users, however. They have also provided a legal basis to this question instead of the Safeguard: though unexpected, this is not incorrect as the excerpt stipulation could be interpreted as a legal basis.

Question 6: The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances.



Information Assets

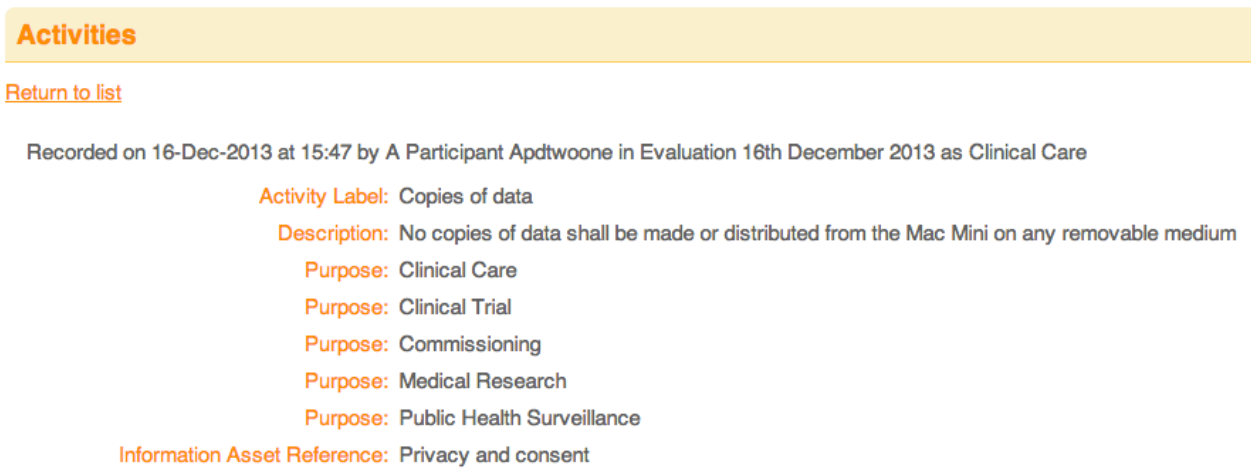
Return to list

Recorded on 16-Dec-2013 at 15:44 by A Participant Apdtwoone in Evaluation 16th December 2013 as Clinical Care

Label:	Mac Mini
Description:	The Mac Mini shall not be networked using the ethernet connection etc etc
Asset Type:	Software
Metadata Format:	Unknown
Hardware Storage:	Laptop
Hardware Storage:	Remote Facility / Cloud

Response Review: The participant has supplied an Information Asset only for this question, with incorrect details provided for Asset Type and two Hardware Storage instances. They have not supplied a Safeguard or another Information Asset in the form of a PDA.

Question 7: No copies of data shall be made or distributed from the Mac Mini on any removable medium.



Activities

Return to list

Recorded on 16-Dec-2013 at 15:47 by A Participant Apdtwoone in Evaluation 16th December 2013 as Clinical Care

Activity Label:	Copies of data
Description:	No copies of data shall be made or distributed from the Mac Mini on any removable medium
Purpose:	Clinical Care
Purpose:	Clinical Trial
Purpose:	Commissioning
Purpose:	Medical Research
Purpose:	Public Health Surveillance
Information Asset Reference:	Privacy and consent

Response Review: The participant has supplied an Activity only for this question instead of a Safeguard. This is not expected though not incorrect, as the excerpt

describes an activity of sharing information. They should have included the CD ROM information Asset, however.

Participant APD_00016

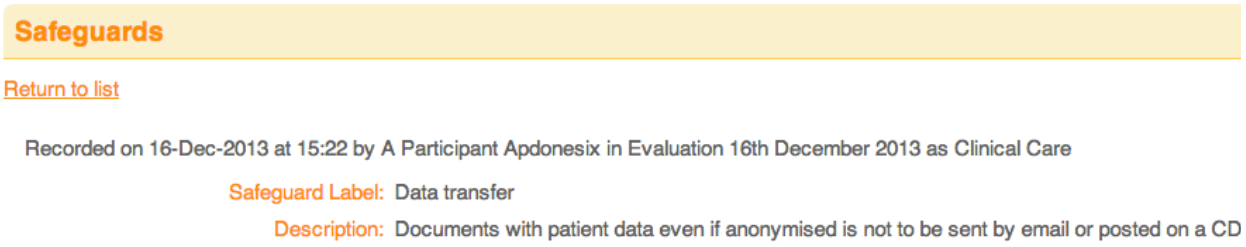
Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.

Response Review: The participant has specified the Information Asset details.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.

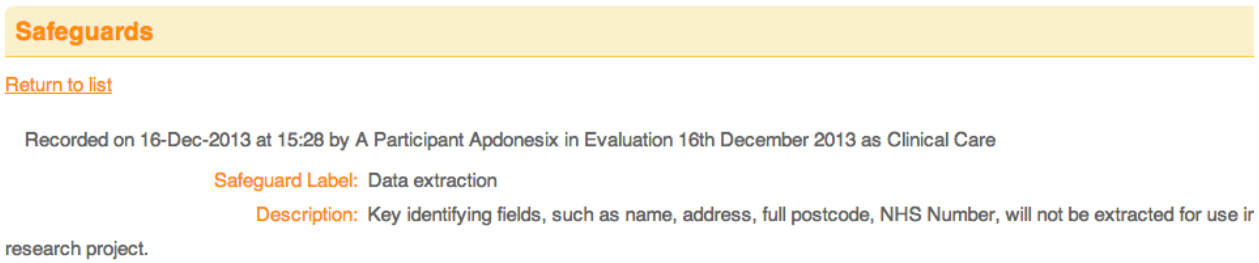
Response Review: The provided the correct details for this question.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.



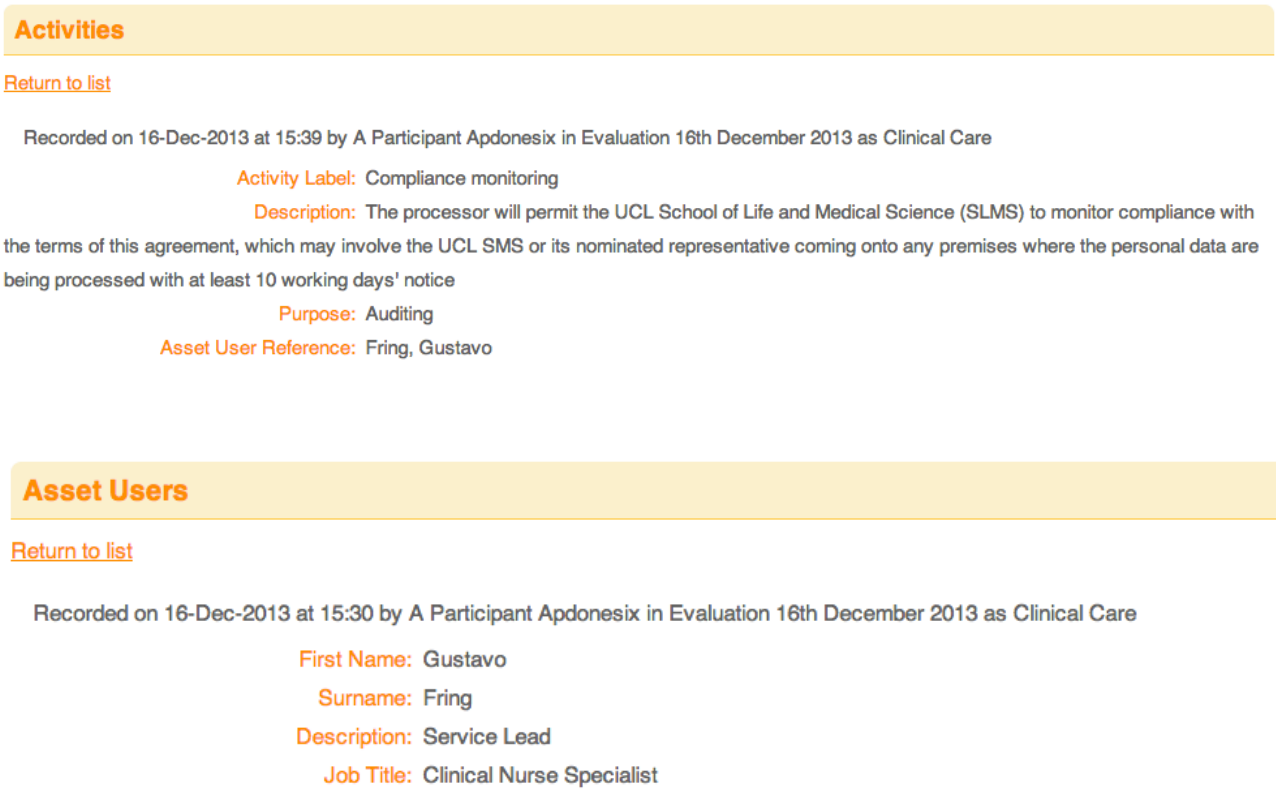
Response Review: The participant has specified a Safeguard with correct details, though there is no reference to the CD ROM Information Asset, and no Control . Additionally, the Activity of Sharing Information has also not been provided.

Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.



Response Review: The participant has specified a Safeguard with a Label and Description, but no Control and no reference to the identifying fields that were added. There is also no reference to the Activity of sharing information, which has not been provided, nor a reference to Research.

Question 5: Audit: The Processor will permit the UCL SLMS to monitor compliance with the terms of this agreement, which may involve the UCL SLMS or its nominated representative coming onto any premises where the personal data are being processed with at least 10 working days' notice.



Activities

Return to list

Recorded on 16-Dec-2013 at 15:39 by A Participant Apdonesix in Evaluation 16th December 2013 as Clinical Care

Activity Label: Compliance monitoring

Description: The processor will permit the UCL School of Life and Medical Science (SLMS) to monitor compliance with the terms of this agreement, which may involve the UCL SMS or its nominated representative coming onto any premises where the personal data are being processed with at least 10 working days' notice

Purpose: Auditing

Asset User Reference: Fring, Gustavo

Asset Users

Return to list

Recorded on 16-Dec-2013 at 15:30 by A Participant Apdonesix in Evaluation 16th December 2013 as Clinical Care

First Name: Gustavo

Surname: Fring

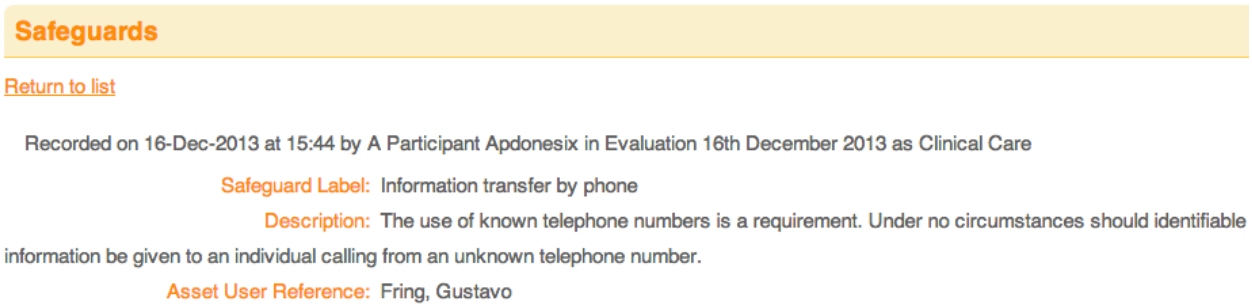
Description: Service Lead

Job Title: Clinical Nurse Specialist

Response Review: As with participant APD_00018, this participant has specified an Activity and an Asset User as expected with sufficient detail. They have applied their professional experience to the details. They have not, however provided a Safeguard, opting instead to place the details of the policy in the Description of the Activity. This approach is not necessarily correct as the stipulations for allowing

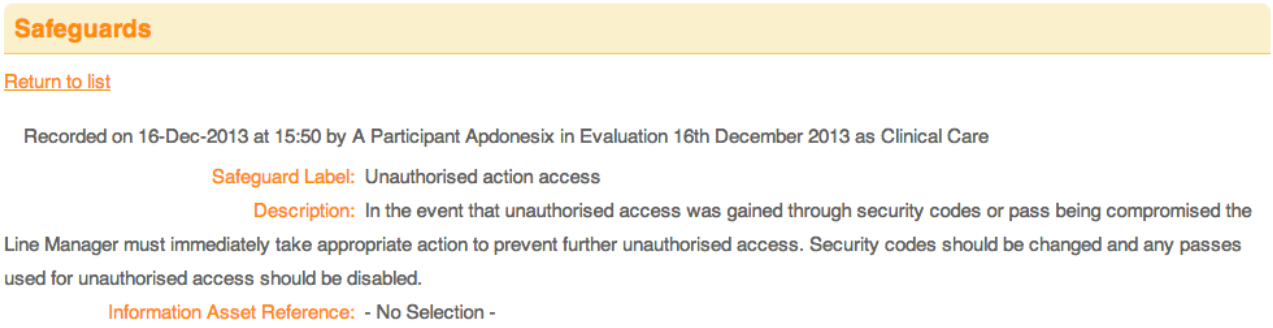
the Audit Activity should be placed in a Safeguard. It is nevertheless arguable that these stipulations could form part of the detail of the Activity and that would be sufficient for users to access the details and behave appropriately, if they knew to look at the detail of the Audit Activity. The Activity details are correct, and include a reference to the Asset User. The Asset User details have had a number of details inferred: Having the description as a Service Lead and Job Title as a Clinical Nurse Specialist is not unreasonable, given that the participant has inferred Audit as a Compliance Monitoring Activity.

Question 6: The use of known telephone numbers is a requirement. Under no circumstances should personally identifiable information be given to an individual calling from an unknown telephone number.



Response Review: The participant has correctly specified a Safeguard, with reasonable details, though they have not provided a Control. They have incorrectly applied this to a specific Asset User. There is also no reference to an Information Asset for known telephone numbers, and they have not supplied an Information Asset for known telephone numbers.

Question 7: In the event that unauthorised access was gained through security codes or passes being compromised the Line Manager must immediately take appropriate action to prevent further unauthorised access. Security codes should be changed and any passes used for unauthorised access should be disabled.



Response Review: The participant has correctly specified a minimal Safeguard with reasonable details, though a No Selection for the Information Asset reference and they have not added a Control. They did not add the Information Assets for Security Codes and Passes.

Experiment 2

Participant APD_00017

General Comments:

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

Looks like there is a data sharing agreement in place; although not clear if data sharing is permitted. If it were then cannot send information via email; even if there are no clear identifiers.

Response Review: The expected response was that anonymised data could be released, but not shared via email or CD, based upon the stipulations authored by Participant APD_00021 in Questions three and four. The question was answered correctly with two responses, both of which were correct.

Question 2: You have received a telephone request for names and addresses of three participants in a research trial. How would you respond?

Response to Question 2:

Would need to know and be able to identify the telephone number that the individual is calling from. Before you are able to identify the person and location they are calling from NO information can be exchanged.

Response Review: This question expected that the participant would use the details added in questions five and six and they made two responses, both of which were correct and expected.

Question 3: A team member from UCL CLMS has asked to come and inspect your compliance with a data sharing agreement you hold with them within the next week. How would you respond to their request?

Response to Question 3:

The participant chose option 3: *I would agree and offer dates ten or more days later.* This was the correct response.

Policy states that SLMS Member is permitted to monitor compliance on premises and this can be granted (i.e. RIV) within a 10 day period of the request.

Response Review: Participant APD_00017 made the correct response and provided the correct justification for it. They specified that they used the Activity details to answer the question correctly, which, in this case, negated the need for the Safeguard that participant APD_00021 did not provide.

Question 4: You have received a report of an unauthorised access to an information asset. How do you respond to this?

Response to Question 4:

Contact and inform appropriate line manager in order to prevent further unauthorised access. If I am the line manager.(or person who can do) then change security codes / passes used for the unauthorised access.

Response Review: Participant APD_00014 made two responses, both of which were correct.

Participant APD_00021

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

keibi Legal basis for inclusion of data, compliance monitoring, Safeguards

Even if anonymised should not be sent by email or posted on a CD.

Activity - question whether prior consent given by patients - if yes then in Safeguards NOT to be sent by email or put on CD. Sent by secure, confidential means

Response Review: The expected response was that anonymised data could be released, but not shared via email or CD, based upon the stipulations authored by Participant APD_00017 in Questions three and four. The question was answered correctly with three responses, all of which were correct.

Question 2: You have received a telephone request for names and addresses of three participants in a research trial. How would you respond?

Response to Question 2:

Need request in writing before responding giving full details of

Trial - including risk and benefits
Need request in writing from P.I. of trial
Need to see patient information sheet and consent form before agreeing to share information.

Response Review: This question expected that the participant would use the details added in questions five and six and they made three responses, none of which were expected: they applied the participant's expertise but did not use the details used in keibi.

Question 3: A team member from UCL CLMS has asked to come and inspect your compliance with a data sharing agreement you hold with them within the next week. How would you respond to their request?

Response to Question 3:

The participant chose option 3: *I would agree and offer dates ten or more days later.* This was the correct response.

keibi under Activity Label - this specified the appropriate details.

Response Review: Participant APD_00021 made the correct response and provided the correct reference for it. They specified that they used the Activity details to answer the question correctly, which, in this case, negated the need for the Safeguard that participant APD_00017 did not provide.

Question 4: You have received a report of an unauthorised access to an information asset. How do you respond to this?

Response to Question 4:

keibi Safeguards - in the event of unauthorised access line manager will take appropriate action to prevent further unauthorised access. Security code to be changed and passes dealt with.

Response Review: Participant APD_00021 made three responses, all of which were correct.

Participant APD_00016
General Comments:

Basis the says NIGB Exemption (now HSCIC, so better to use Section 251)

On Activities - error message 'supplied identifier is not valid for the given property' but no indication of what it refers to.

Information Asset reference - too granular

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

Cannot agree to request (from Safeguards).

Unsure how to proceed - if it is just a list of BP readings? then ok.

Response Review: This question expected that the participant would use the details added in questions three and four. There are three responses: the first response stated that they could not agree to the request. The second response said that they weren't sure how to proceed. The final response was for the participant suggested that transferring as requested might be permissible if it was only the blood pressure results.

Question 2: You have been asked to transfer data from a Mac Mini holding a research data repository to a colleague using a USB key. How would you respond?

Response to Question 2:

Cannot agree to request

keibi - Safeguards: no copies of data from Mac Mini or any removable medium.

If anonymised ? ok on USB key?

Response Review: This question expected that the participant would use the details added in questions six and seven by Participant APD_00017 and refers to the expected Safeguard. There is a single response, which is correct. There was another response that suggests using a USB key if the data is anonymised - this shows that the participant felt that there was insufficient guidance by offering an alternative.

Question 3: A colleague has requested that you network the Mac Mini to perform a critical software update. How would you respond?

Response to Question 3:

Correct answer: "I would speak to the policy authors and ask them whether this is an acceptable case for breaching the policy, or how they would proceed." - But "critical" software update - ? sufficient.

Justification for breach of policy - ? potential consequences if software update not permitted. - Further advice sought and needed.

Response Review: This question expected that the participant would use the details added in question six by Participant APD_00015. The response is correct - the participant questioned the stipulation in keibi and chose the option that did not compromise good practice. Two correct responses were made.

Question 4: A research colleague has asked for some patient data. They say that it can be anonymised and that they would be able to receive it via email, on a posted CD or they will be happy to collect it on an encrypted USB key. How do you respond?

Response to Question 4:

Mac Mini - Removable Medium - Rules out CD / USB Key.

Patient consent required prior to data sharing

Cannot send data by email even if anonymised.

How else would one share with collaborators?

However, does "data" relate to full data set?

Therefore can anonymised data set be placed on USB key?

Response Review: This question expected that the participant would use the details added in question four, five, six and seven by Participant APD_00017. There are three responses, all of which are correct - the participant understood the pertinent details and responded correctly, referring to the appropriate Safeguards. The participant also proposes alternatives - this exceeds expectations as they are wondering how they might achieve the request.

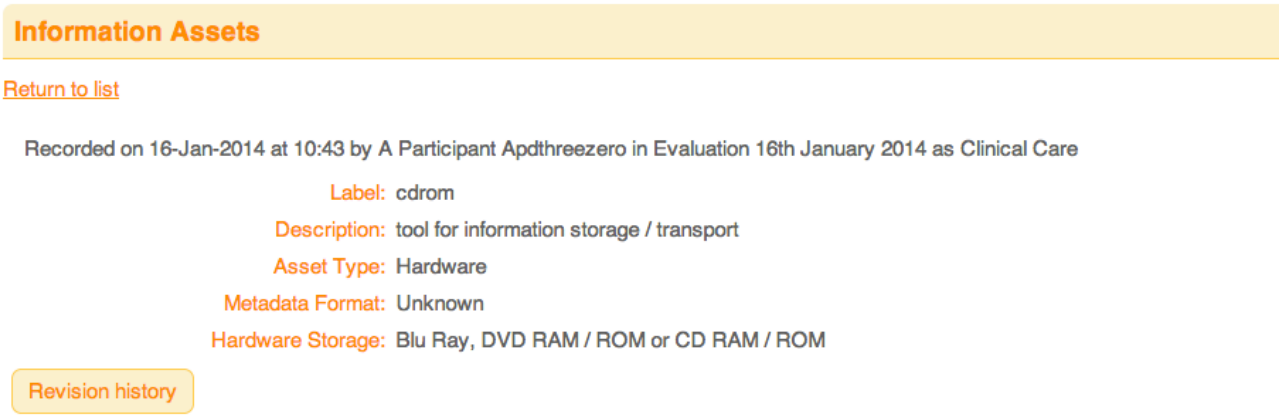
Evaluation Session 16th January 2014

Participant APD_00030

The responses from participant APD_00030 are provided below.

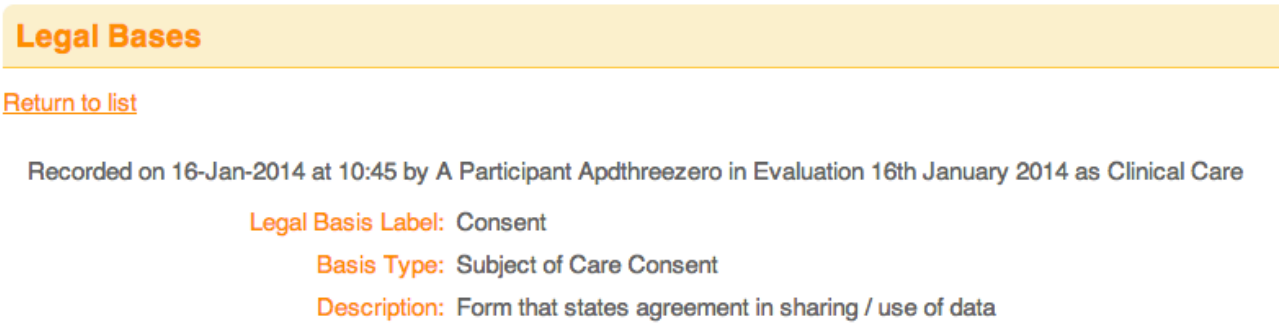
Experiment One

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.



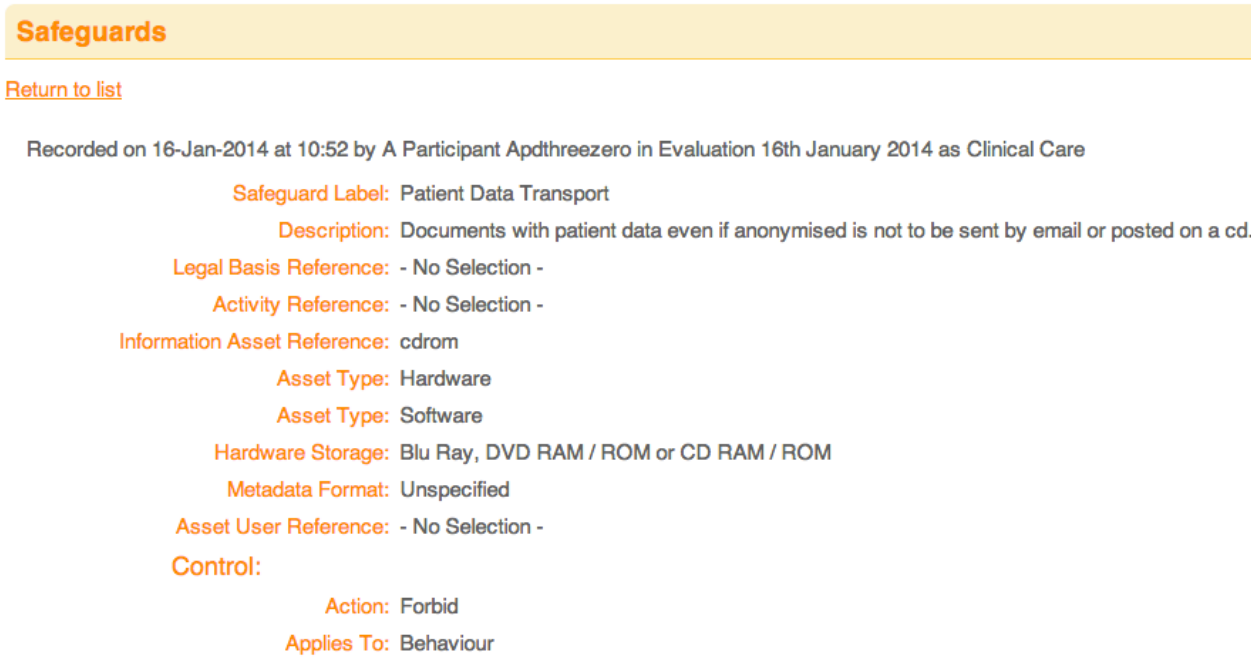
Response Review: the participant was able to use the features of *keibi* to add details that represented a CD ROM Information Asset.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.



Response Review: the participant has correctly added the legal basis of Consent.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.



Safeguards

Return to list

Recorded on 16-Jan-2014 at 10:52 by A Participant Apdthreezero in Evaluation 16th January 2014 as Clinical Care

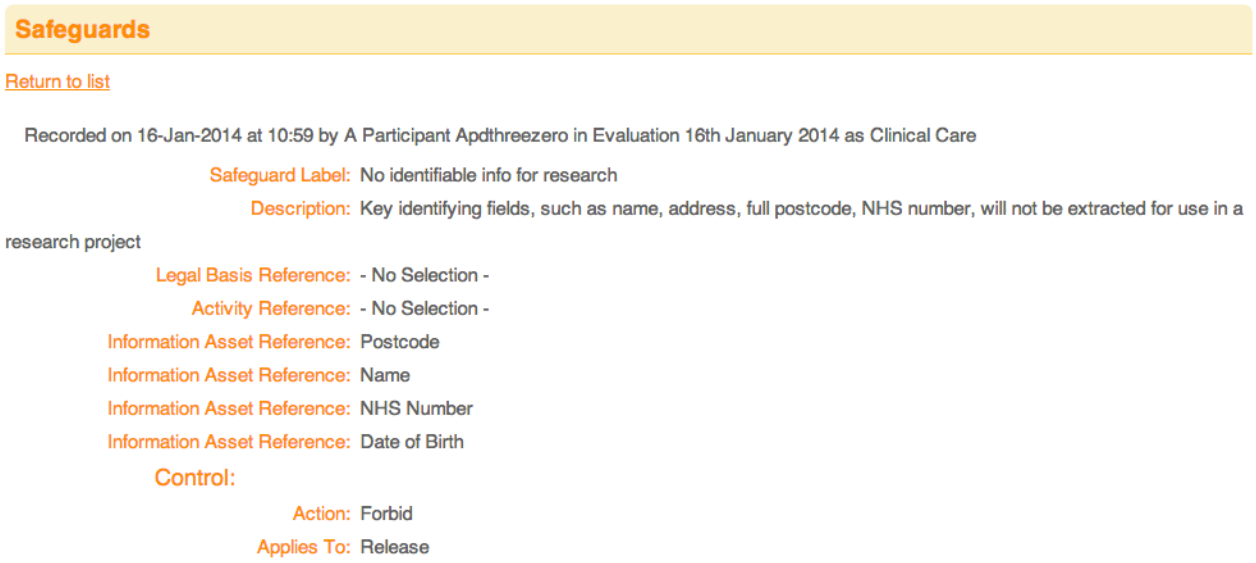
Safeguard Label:	Patient Data Transport
Description:	Documents with patient data even if anonymised is not to be sent by email or posted on a cd.
Legal Basis Reference:	- No Selection -
Activity Reference:	- No Selection -
Information Asset Reference:	cdrom
Asset Type:	Hardware
Asset Type:	Software
Hardware Storage:	Blu Ray, DVD RAM / ROM or CD RAM / ROM
Metadata Format:	Unspecified
Asset User Reference:	- No Selection -

Control:

Action:	Forbid
Applies To:	Behaviour

Response Review: the participant has correctly added a Safeguard that describes the required information. Both Legal Basis and Activity references have no selection, and they have not provided the Sharing Data Activity. The other details are reasonable. They correctly refer to the CD ROM Information Asset. They have also not provided an email Activity.

Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.



Safeguards

Return to list

Recorded on 16-Jan-2014 at 10:59 by A Participant Apdthreezero in Evaluation 16th January 2014 as Clinical Care

Safeguard Label: No identifiable info for research

Description: Key identifying fields, such as name, address, full postcode, NHS number, will not be extracted for use in a research project

Legal Basis Reference: - No Selection -

Activity Reference: - No Selection -

Information Asset Reference: Postcode

Information Asset Reference: Name

Information Asset Reference: NHS Number

Information Asset Reference: Date of Birth

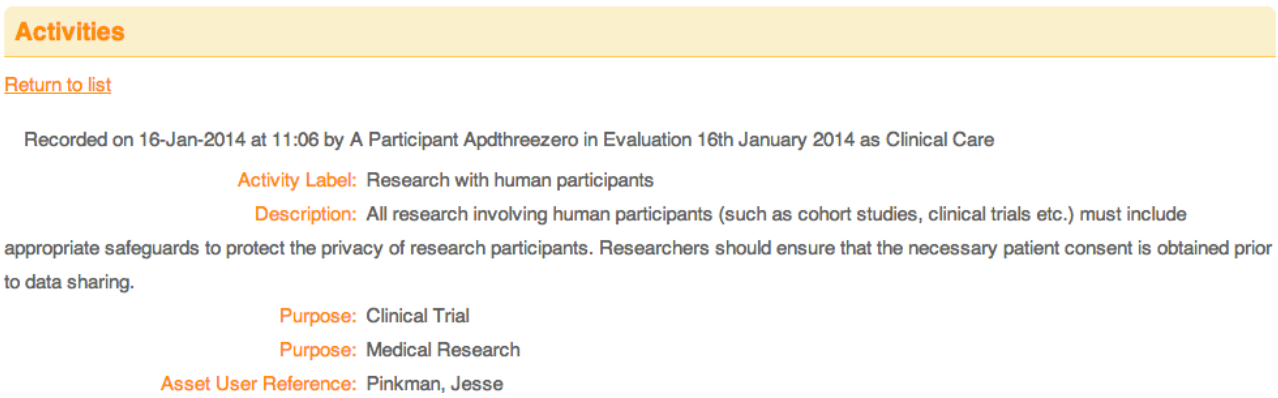
Control:

Action: Forbid

Applies To: Release

Response Review: the participant has correctly added the Safeguard in this case for the most part, though they have not provided any selections for Legal Basis or Activity references. The correct Control has also been added. The participant was expected to add an Activity for describing Research but has not done this.

Question 5: All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.



Activities

Return to list

Recorded on 16-Jan-2014 at 11:06 by A Participant Apdthreezero in Evaluation 16th January 2014 as Clinical Care

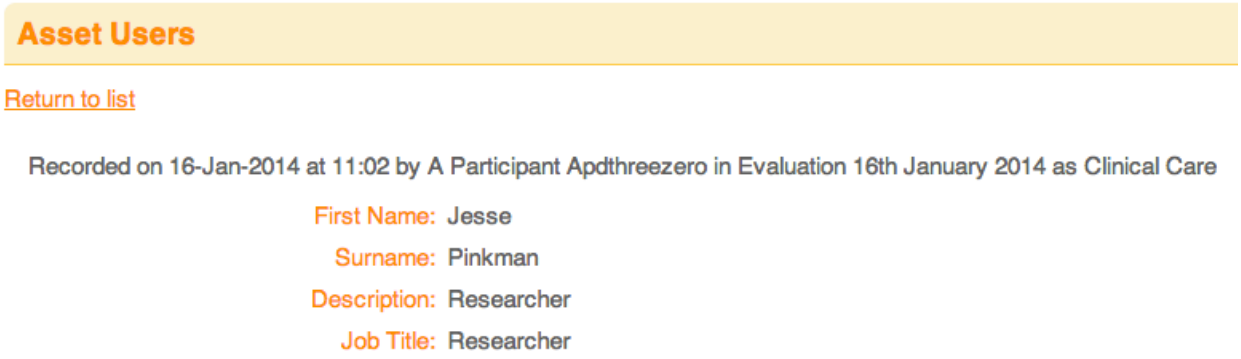
Activity Label: Research with human participants

Description: All research involving human participants (such as cohort studies, clinical trials etc.) must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.

Purpose: Clinical Trial

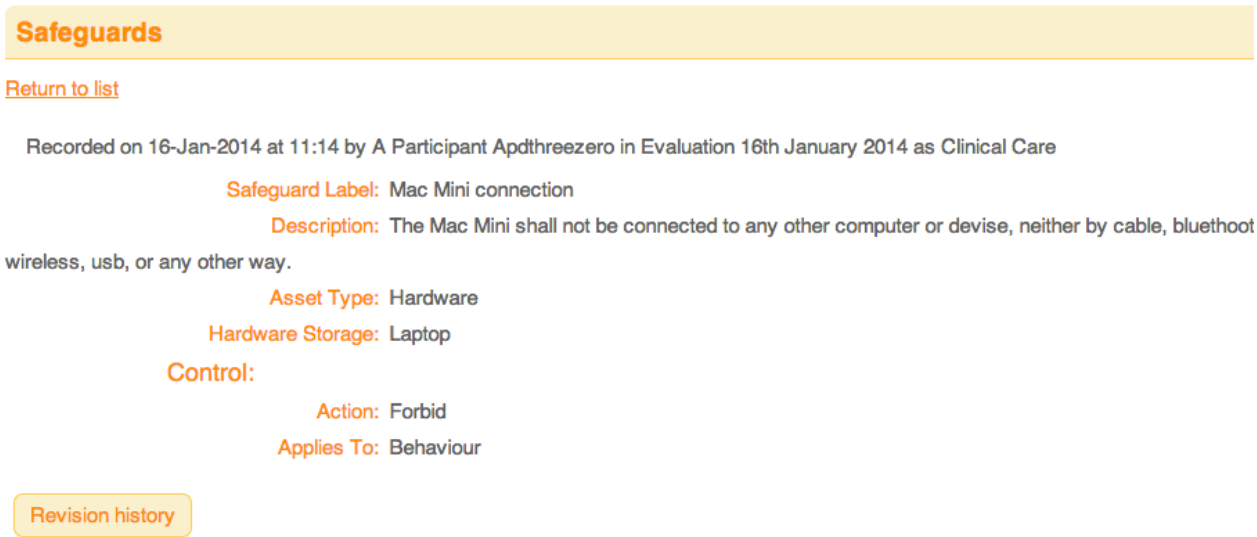
Purpose: Medical Research

Asset User Reference: Pinkman, Jesse



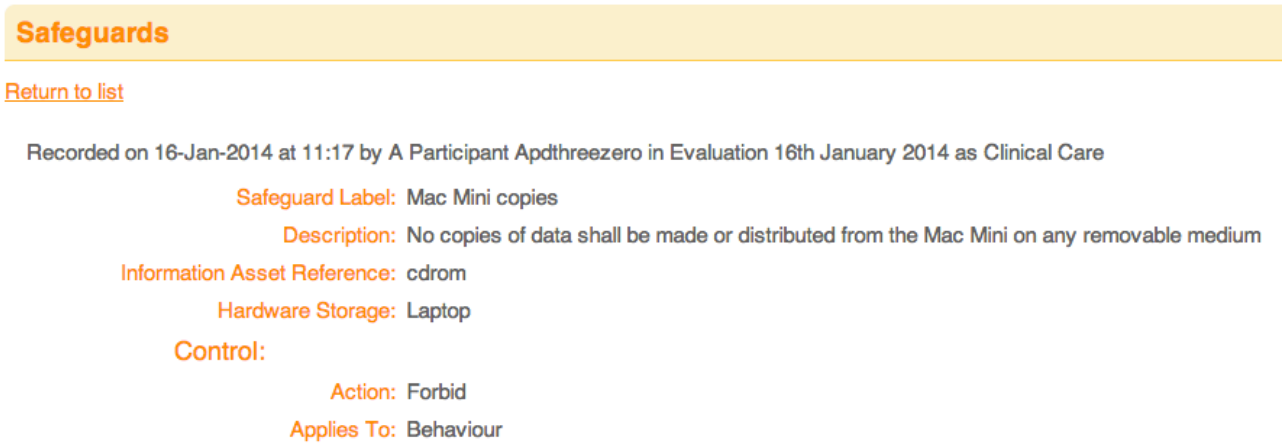
Response Review: the participant has correctly added the Asset User and a Research Activity, with the correct reference to the Asset User, though this might have been added in the previous question. The details are all reasonable, however they have not added a Safeguard in this case.

Question 6: The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances.



Response Review: The participant has specified the expected Safeguard correctly for the most part, tough Hardware Storage of Laptop is incorrect. They have also not specified an Information Asset to describe the Mac Mini or PDA.

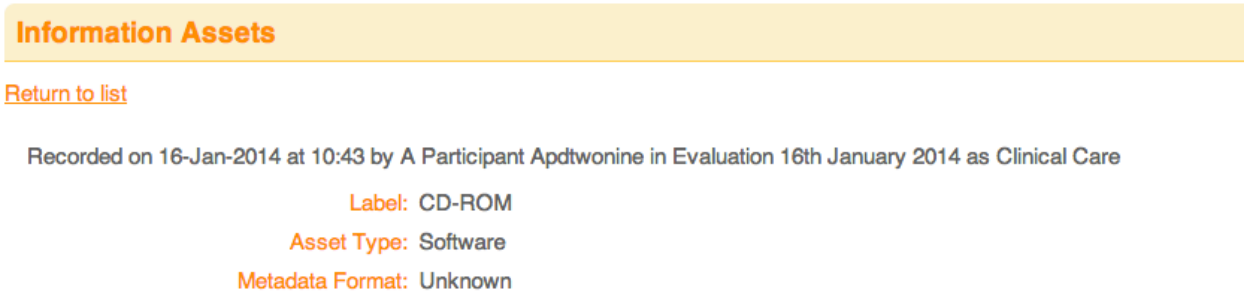
Question 7: No copies of data shall be made or distributed from the Mac Mini on any removable medium.



Response Review: The Safeguard for preventing copies being made from the Mac Mini is also correct, has correct details, except for the Hardware Storage of Laptop.

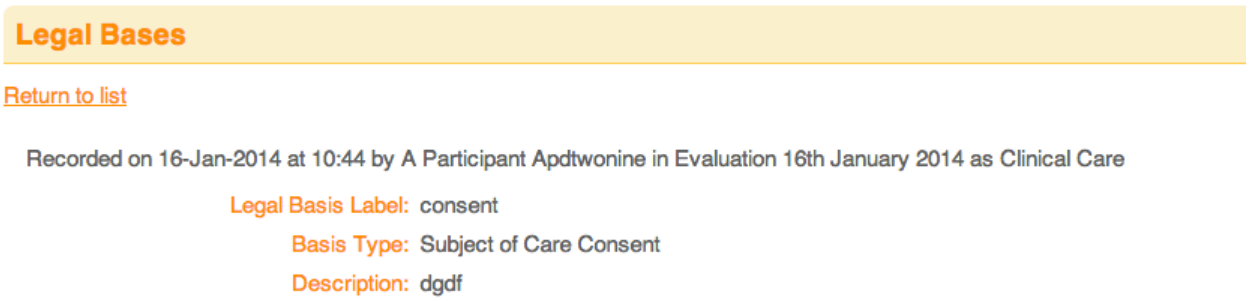
Participant APD_00029

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.



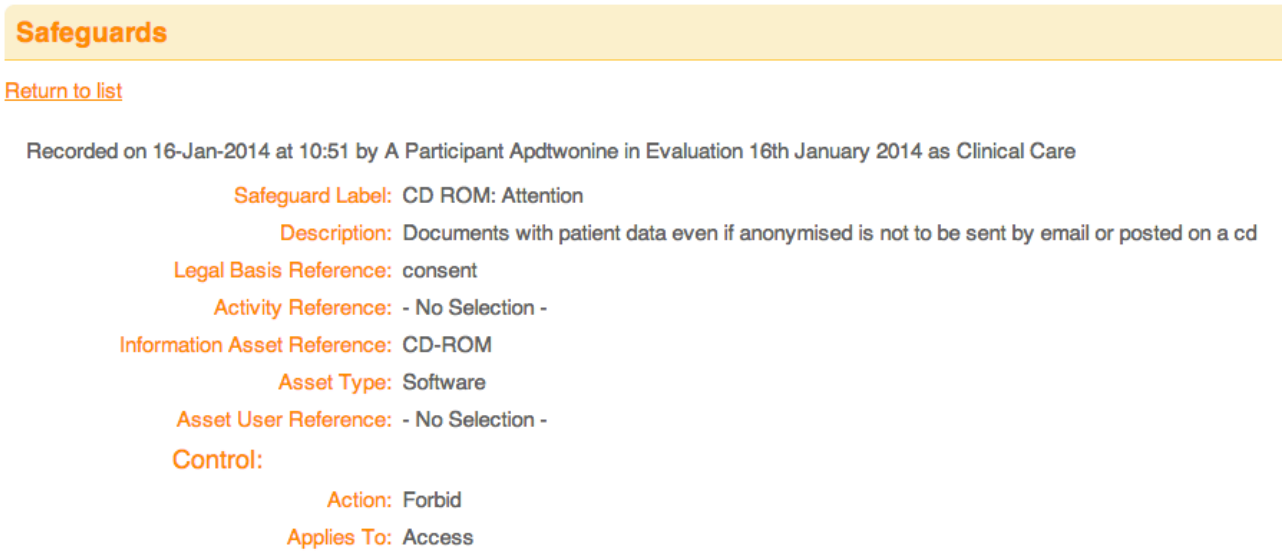
Response Review: The participant has specified the Information Asset details, though the Asset Type is not software.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.



Response Review: The participant added the details correctly, with the exception of the description field.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.



Safeguards

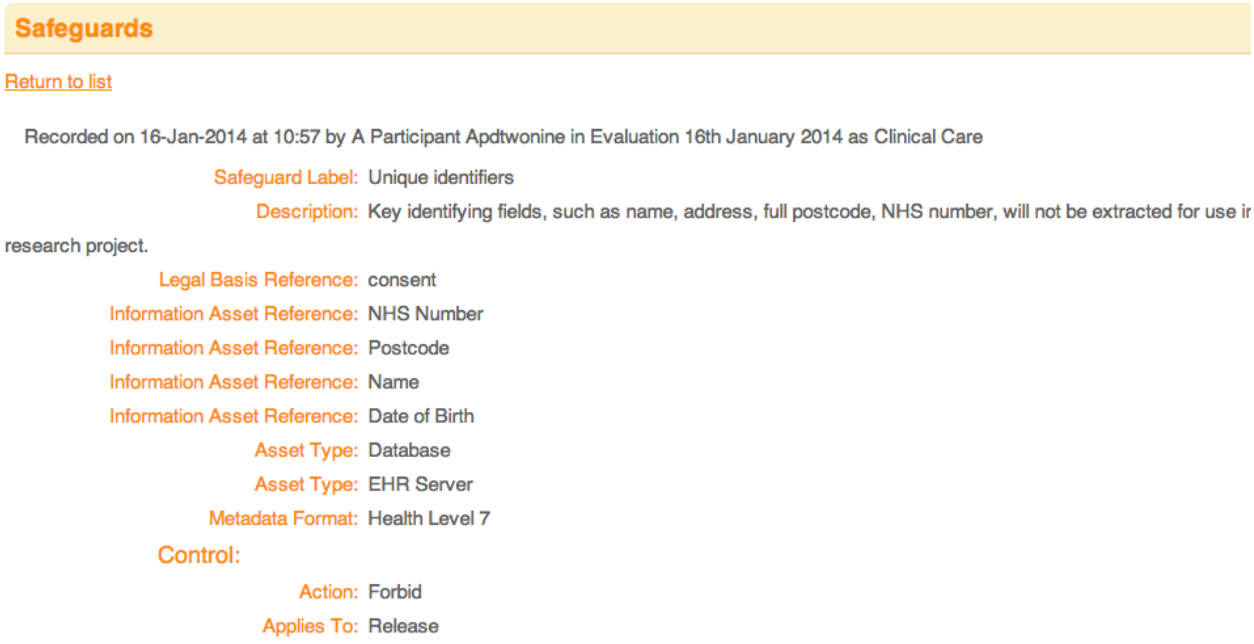
Return to list

Recorded on 16-Jan-2014 at 10:51 by A Participant Apdtwonine in Evaluation 16th January 2014 as Clinical Care

Safeguard Label: CD ROM: Attention
Description: Documents with patient data even if anonymised is not to be sent by email or posted on a cd
Legal Basis Reference: consent
Activity Reference: - No Selection -
Information Asset Reference: CD-ROM
Asset Type: Software
Asset User Reference: - No Selection -
Control:
Action: Forbid
Applies To: Access

Response Review: The participant has specified a Safeguard with correct details, though there is no reference to the sharing information Activity. The Control detail for Applies to is also incorrect. There is a no selection for Asset User Reference and the Asset Type of Software is incorrect. Additionally, the Activity of Sharing Information has also not been provided.

Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.



Safeguards

Return to list

Recorded on 16-Jan-2014 at 10:57 by A Participant Apdtwonine in Evaluation 16th January 2014 as Clinical Care

Safeguard Label: Unique identifiers

Description: Key identifying fields, such as name, address, full postcode, NHS number, will not be extracted for use ir research project.

Legal Basis Reference: consent

Information Asset Reference: NHS Number

Information Asset Reference: Postcode

Information Asset Reference: Name

Information Asset Reference: Date of Birth

Asset Type: Database

Asset Type: EHR Server

Metadata Format: Health Level 7

Control:

Action: Forbid

Applies To: Release

Response Review: The participant has specified a Safeguard with correct details, with the exception of the Asset Types and Metadata format. They have not specified an Activity for Research.

Question 5: Audit: The Processor will permit the UCL SLMS to monitor compliance with the terms of this agreement, which may involve the UCL SLMS or its nominated representative coming onto any premises where the personal data are being processed with at least 10 working days' notice.

Safeguards

Return to list

Recorded on 16-Jan-2014 at 11:09 by A Participant Apdtwonine in Evaluation 16th January 2014 as Clinical Care

Description: Audit: The Processor will permit the ucl alms to monitor compliance with the terms of this agreement, which may involve the ucl SLMS or its nominated representative coming onto any premises where the personal data are being processed with at least 10 working days notice

Legal Basis Reference: consent
Activity Reference: Processor of personal data
Information Asset Reference: Postcode
Information Asset Reference: Name
Information Asset Reference: NHS Number
Information Asset Reference: Date of Birth
Asset Type: Database
Asset Type: EHR Server
Asset User Reference: White, Walter

Asset Users

Return to list

Recorded on 16-Jan-2014 at 11:04 by A Participant Apdtwonine in Evaluation 16th January 2014 as Clinical Care

First Name: Walter
Surname: White
Description: Nominated UCL representative
Job Title: UCL Representative
Affiliated Organisation: UCL SLMS
Responsibility: Monitor compliance with terms of agreement
Role: UCL representative

Activities

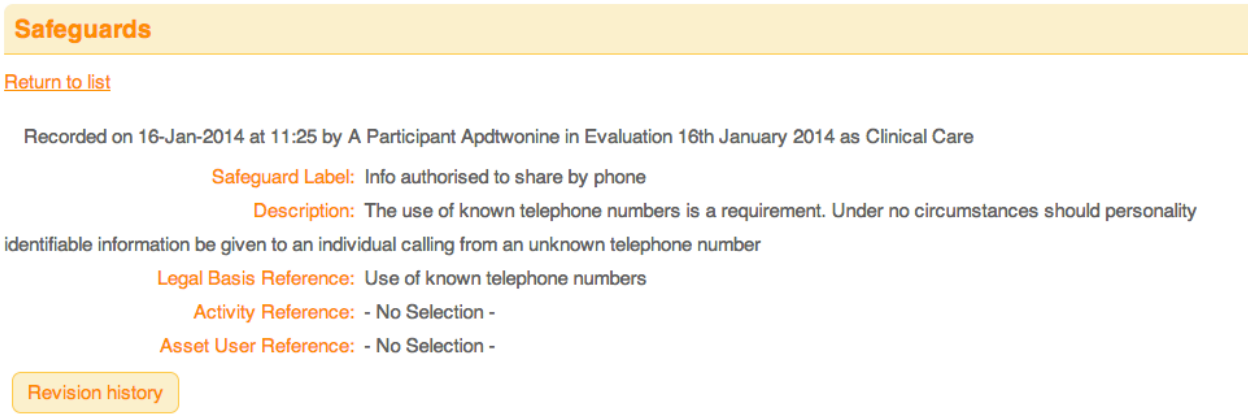
Return to list

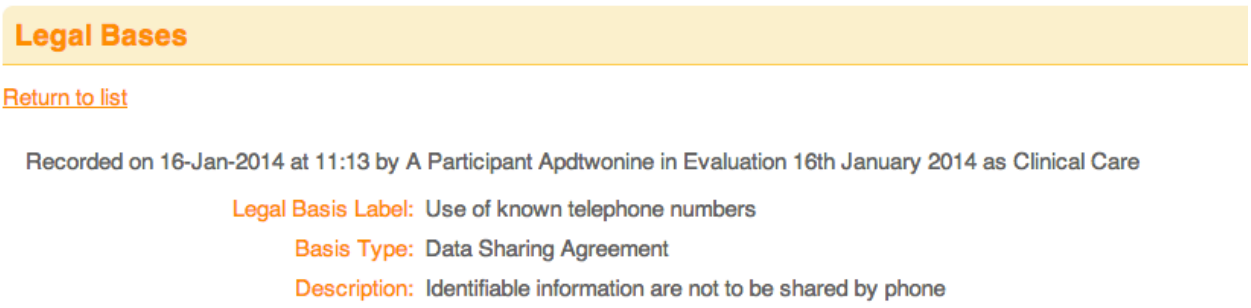
Recorded on 16-Jan-2014 at 11:05 by A Participant Apdtwonine in Evaluation 16th January 2014 as Clinical Care

Activity Label: Processor of personal data
Purpose: Auditing
Information Asset Reference: Postcode
Information Asset Reference: Name
Information Asset Reference: NHS Number
Information Asset Reference: Date of Birth
Asset User Reference: White, Walter

Response Review: The participant has specified a Safeguard with no Label and a correct Description, as well as Asset User and Activity reference. No Control has been specified, however. The other details are incorrect. They have provided correct details for the Asset User, and a Processes Personal Data Activity, which is unexpected and also not applicable here, though it contains correct details for an Audit Activity.

Question 6: The use of known telephone numbers is a requirement. Under no circumstances should personally identifiable information be given to an individual calling from an unknown telephone number.





Response Review: The participant has correctly specified a Safeguard, with a reasonable label and Description, though the Activity and Asset User references are incorrect. They have opted to specify a Legal Basis for using Telephone Numbers, which is not incorrect, though the Legal Basis description is incorrect. They have also not specified the known Telephone Number Information Asset.

Question 7: In the event that unauthorised access was gained through security codes or passes being compromised the Line Manager must immediately take appropriate action to prevent further unauthorised access. Security codes should be changed and any passes used for unauthorised access should be disabled.

Safeguards

Return to list

Recorded on 16-Jan-2014 at 11:24 by A Participant Apdtwonine in Evaluation 16th January 2014 as Clinical Care

Description: In the event that unauthorised access was gained through security codes or passes being compromised the line manager must immediately take appropriate action to prevent further unauthorised access. security codes should be changed and any passes used for unauthorised access should be disabled.

Legal Basis Reference: Unauthorised access
Activity Reference: - No Selection -
Information Asset Reference: - No Selection -
Asset Type: Software
Asset Type: Database
Asset Type: EHR Server
Asset User Reference: James, Jesse

Asset Users

Return to list

Recorded on 16-Jan-2014 at 11:21 by A Participant Apdtwonine in Evaluation 16th January 2014 as Clinical Care

First Name: Jesse
Surname: James
Description: Responsible for action when unauthorised access gained
Job Title: Line manager
Affiliated Organisation: UCL SLMS
Responsibility: change security codes
Responsibility: disable unauthorised passes

Response Review: The participant has correctly specified a Safeguard with reasonable details, though a No Selection for the Information Asset reference and Activity Reference. They did not add the Information Assets for Security Codes and Passes. They have also added another Asset User called Jesse James, for the Line Manager. This is unexpected and not incorrect, containing reasonable details.

Experiment 2

Participant APD_00030

General Comments:

"consumable" as asset type

Unclear on Unique Identifiers

Safeguards |:| with assets?

You already know the assets type, why as again?

"Supplied Identifier is not valid for given property" - mark incorrect fields.

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

I checked the Activities and Safeguards. There is a relevant Safeguard stating that "Patient Data even if anonymised is not to be sent by email." So I would Decline the request as such I suggest I may be able to put requested info on an enclosed secure drive or hand over a CD ROM personally.

Response Review: The expected response was that anonymised data could be released, but not shared via email or CD, based upon the stipulations authored by Participant APD_00029 in Questions three and four. The question was answered correctly with three responses, all of which were correct and provided a reasonable alternative.

Question 2: You have received a telephone request for names and addresses of three participants in a research trial. How would you respond?

Response to Question 2:

There is a Safeguard stating that no personal identifiable information is to shared with people calling from an unknown number. So I would check the number. I guess they mean it should appear on some list. (However it is a little ambiguous: one could say the number is known if it appears on my phone display.)

N.B. it does not matter as identifiable information is not to be shared with researchers (not even for recruitment, apparently [Safeguard 4].

Response Review: This question expected that the participant would use the details added in questions five and six. The participant made five responses, all of which were reasonable and two surpassed expectations: it is true that the Safeguard specified in Question four by Participant APD_00029 does conflict with the Safeguard describing use of telephone numbers. It is also arguable that the original excerpt was vague in what was meant by "known" telephone numbers.

Question 3: A team member from UCL CLMS has asked to come and inspect your compliance with a data sharing agreement you hold with them within the next week. How would you respond to their request?

Response to Question 3:

The participant chose option 4: *I would agree and offer dates ten or more days later.* This was the correct response.

A SLMS representative can expect compliance for checking the premises as long as given 10 days or more notice.

Response Review: Participant APD_00030 made the correct response and provided the correct justification for it.

Question 4: You have received a report of an unauthorised access to an information asset. How do you respond to this?

Response to Question 4:

I would immediately inform my line manager. If I have this authority, I would change passwords (or let IT do it) and/or disable access pass.

Response Review: Participant APD_00030 made three responses, all of which were correct.

Participant APD_00029

General Comments:

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

1) I would initially decline the request, due to the Safeguard policy regulation, according to which patients' data are not to be sent via email, or posted to a CD.
2) I would speak to the policy makers to advice me on how to resolve the issue. If permission was granted, I would make sure data were anonymised and proceed with the query.

Response Review: This question expected that the participant would use the details added in questions three and four. The participant has provided four responses: two of them are expected and correct, whilst the other two exceed expectations and show the participant has supplied their own expertise: speaking to the policy makers and seeking advice on how to proceed was not an expected response, nether was following their advice.

Question 2: You have been asked to transfer data from a Mac Mini holding a research data repository to a colleague using a USB key. How would you respond?

Response to Question 2:

1) According to Safeguard policies 3 and 4, the Mac Mini should not be connected to any other comp. or device, neither by cable, USB…, or any other way. Moreover, cannot distribute any copy of data via Mac Mini.
2) I would ask for the policy makers' permission. If granted,
3) I would make sure that there is no key identifiable info for research and proceed with the query.

Response Review: This question expected that the participant would use the details added in questions six and seven by Participant APD_00018 and refers to the expected Safeguard. There are five responses, two of which are correct. The decision to ask the policy makers' permission to proceed exceeds expectations, and the participant has correctly and unexpectedly applied the stipulations provided in Question 4 to remove key identifying fields.

Question 3: A colleague has requested that you network the Mac Mini to perform a critical software update. How would you respond?

Response to Question 3:

Correct Answer - option 4: I would speak to the policy authors and ask them whether this is an acceptable case for breaching the policy, or how they would proceed.

Response Review: This question expected that the participant would use the details added in question six by Participant APD_00030. The response is correct - the participant felt that they could question the stipulations that had been stored in *keibi*. This suggests a preference to question what is stored in the tool and obey good practice.

Question 4: A research colleague has asked for some patient data. They say that it can be anonymised and that they would be able to receive it via email, on a posted CD or they will be happy to collect it on an encrypted USB key. How do you respond?

Response to Question 4:

1) I would let the policy makers know about the case.
2) I would make sure the data are anonymised.
3) Use an encrypted USB key to finalise the transaction.

Response: No.

Response Review: This question expected that the participant would use the details added in question four, five, six and seven by Participant APD_00030. There are three responses, the first two of which are correct, one of which is unexpected. The final response is incorrect - there is no indication that use of an Encrypted USB is appropriate in this case.

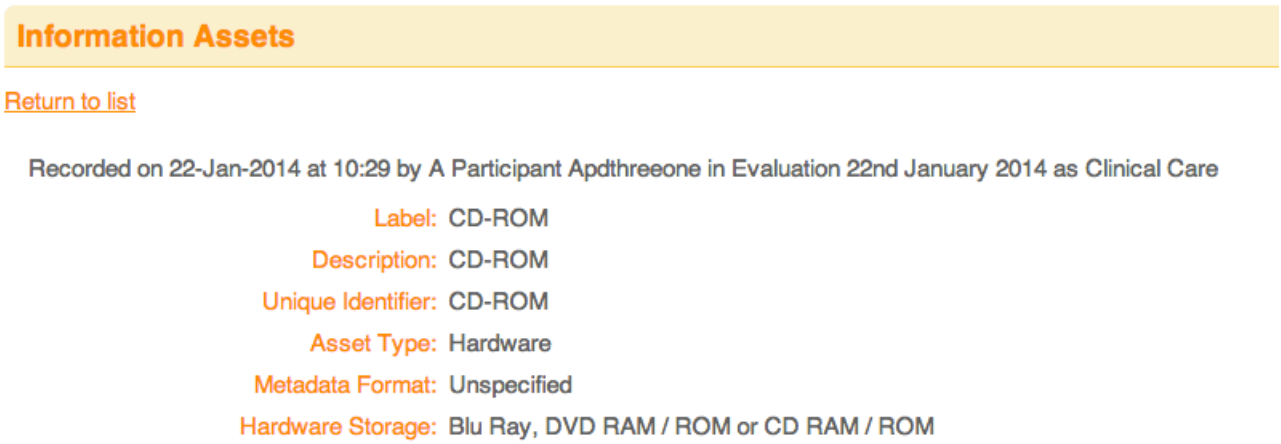
22nd January 2014

Participant APD_00031

The responses from participant APD_00017 are provided below.

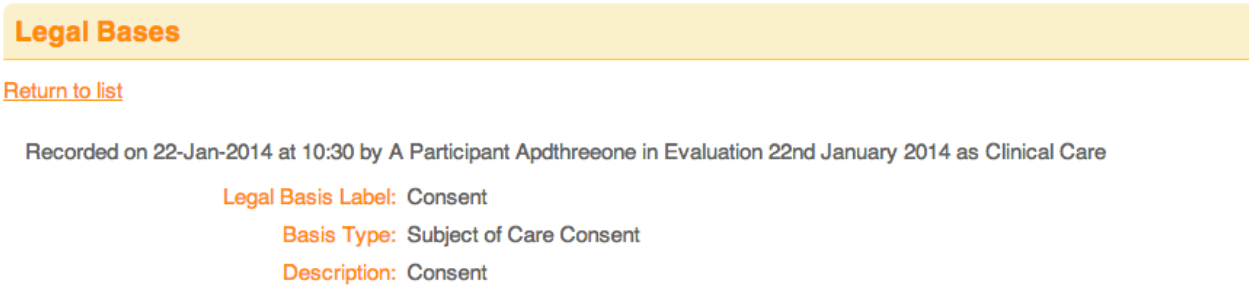
Experiment One

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.



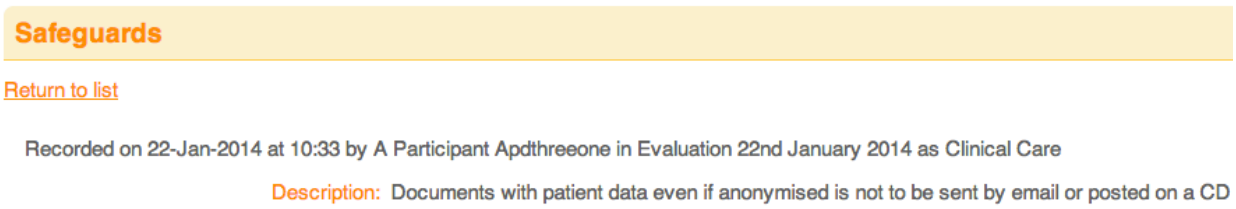
Response Review: the participant was able to use the features of *keibi* to add details that represented a CD ROM Information Asset, Though the Unique Identifier CD ROM is not uniquely identifying.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.



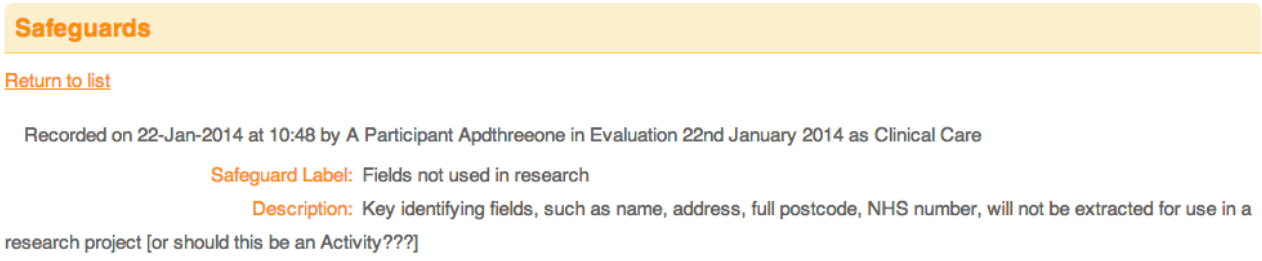
Response Review: the participant has correctly added the legal basis of Consent.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.



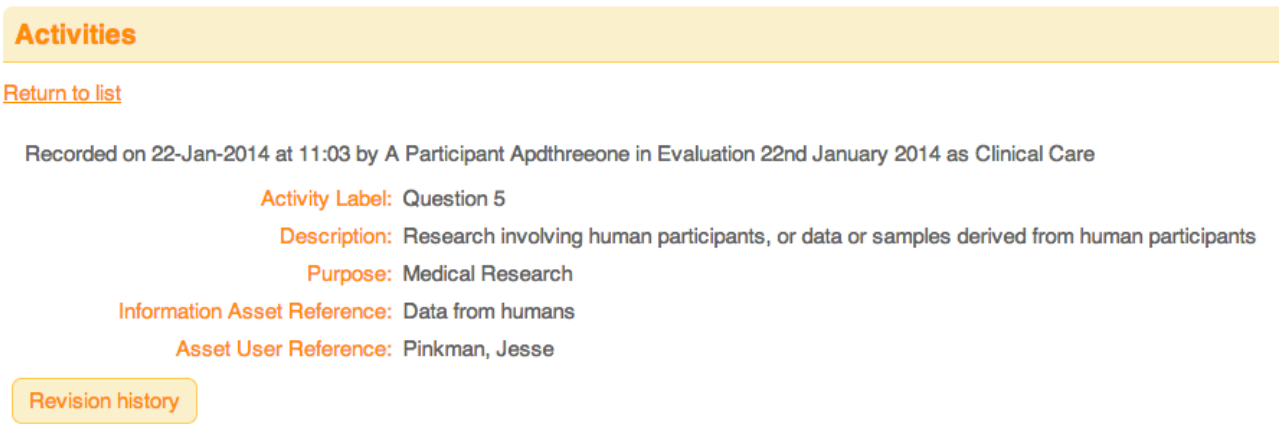
Response Review: the participant has correctly added a Safeguard that describes the a description of the stipulation, but no other details. They have also not added the Activities of email or sharing data.

Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.



Response Review: the participant has correctly added the Safeguard in this case, but has only added a label and description, with no reference to the Information Assets that had been pre-prepared, and no Control. They have not referred to a Research Activity or added it.

Question 5: All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.



Asset Users

Return to list

Recorded on 22-Jan-2014 at 10:52 by A Participant Apdthreeone in Evaluation 22nd January 2014 as Clinical Care

First Name: Jesse
Surname: Pinkman
Description: A fictitious researcher
Job Title: Researcher
Affiliated Organisation: UCL SLMS
Responsibility: Research
Role: Researcher

Information Assets

Return to list

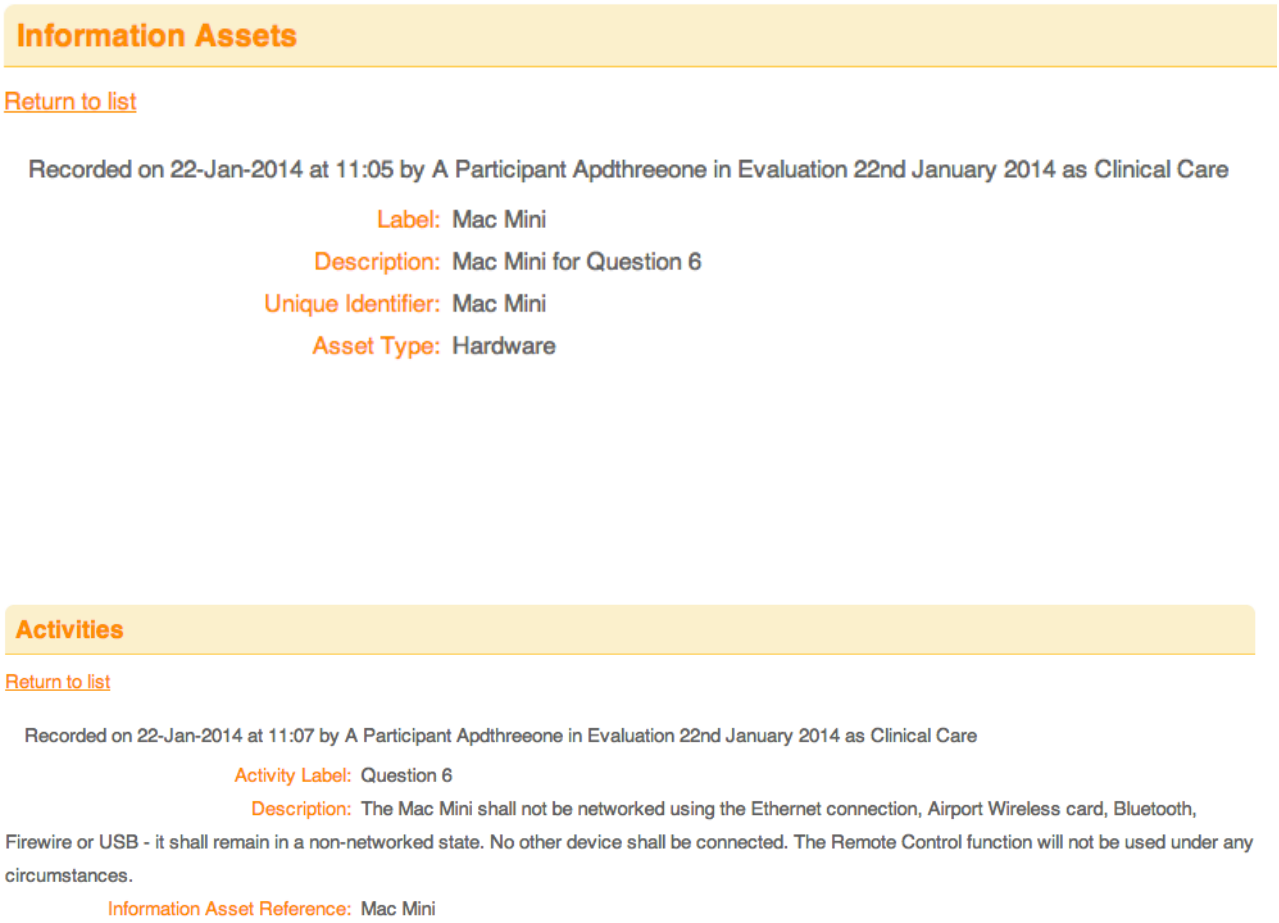
Recorded on 22-Jan-2014 at 11:02 by A Participant Apdthreeone in Evaluation 22nd January 2014 as Clinical Care

Label: Data from humans
Description: Data from research involving human participants

Revision history

Response Review: the participant has added an Activity instead of a Safeguard, but with appropriate detail in most fields, but no Control. The participant has started to label the Compositions according to the Question they are answering as well, which the investigator did not request. They have correctly added a reference to the Asset User and have unexpectedly added a reference to Data from Humans, as well as adding the Information Asset. They have added an Asset User with reasonable details.

Question 6: The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances.



Information Assets

Return to list

Recorded on 22-Jan-2014 at 11:05 by A Participant Apdthreeone in Evaluation 22nd January 2014 as Clinical Care

Label: Mac Mini
Description: Mac Mini for Question 6
Unique Identifier: Mac Mini
Asset Type: Hardware

Activities

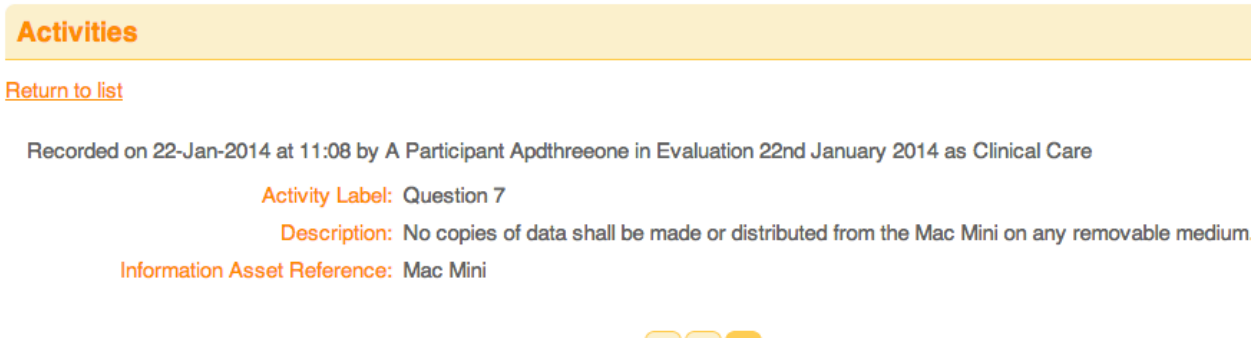
Return to list

Recorded on 22-Jan-2014 at 11:07 by A Participant Apdthreeone in Evaluation 22nd January 2014 as Clinical Care

Activity Label: Question 6
Description: The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device shall be connected. The Remote Control function will not be used under any circumstances.
Information Asset Reference: Mac Mini

Response Review: The participant has specified the expected Information Asset correctly. They have not, however, specified the PDA Information Asset. They have also failed to supply the non networking Safeguard, but have represented this in an Activity with a Correct Reference to the Mac Mini Information Asset.

Question 7: No copies of data shall be made or distributed from the Mac Mini on any removable medium.



Response Review: The Safeguard for sharing data from the Mac Mini is correct, with the correct reference to the Mac Mini Information Asset. There is also the omission of the CD ROM Information Asset reference, and the addition of a misleading Label.

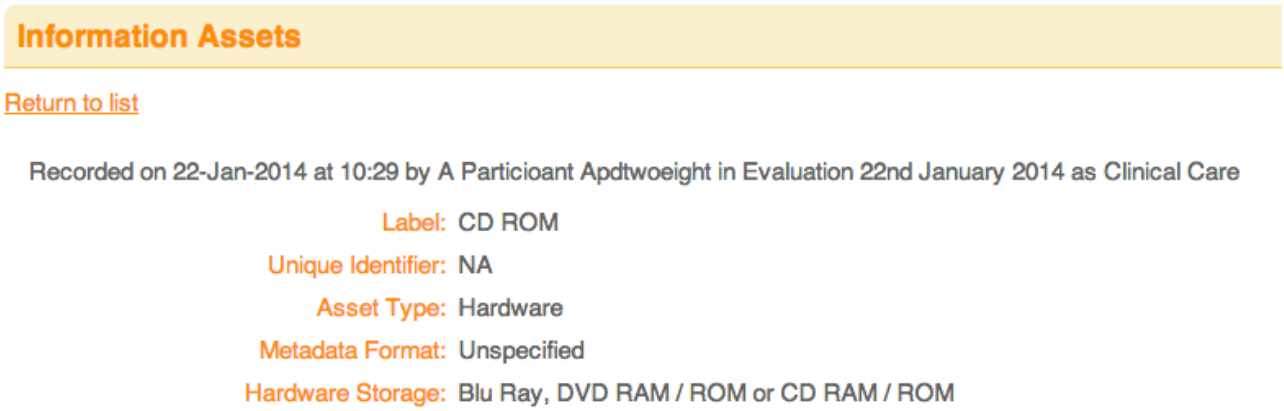
Participant APD_00028

The responses from participant APD_00028 are provided below.

Experiment One

The first experiment asked all participants to author policy excerpts in *keibi*. Section xxx provides details of the questions posed in this experiment, and the excerpts are repeated here for convenience.

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.

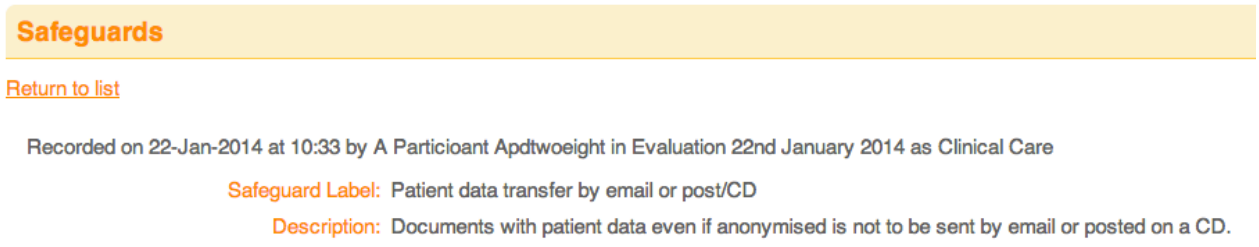


Response Review: the participant was able to use the features of *keibi* to add details that represented a CD ROM Information Asset.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.

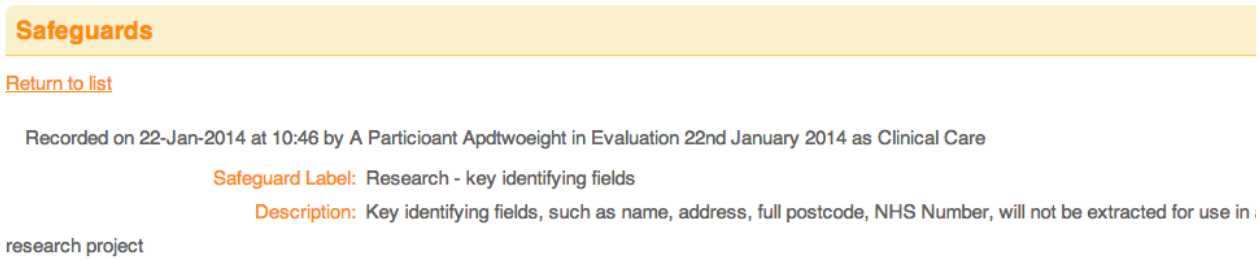
Response Review: the participant has correctly added the legal basis of Consent.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.



Response Review: the participant has supplied an expected Safeguard but with little detail. They have also not provided the sharing information Activity.

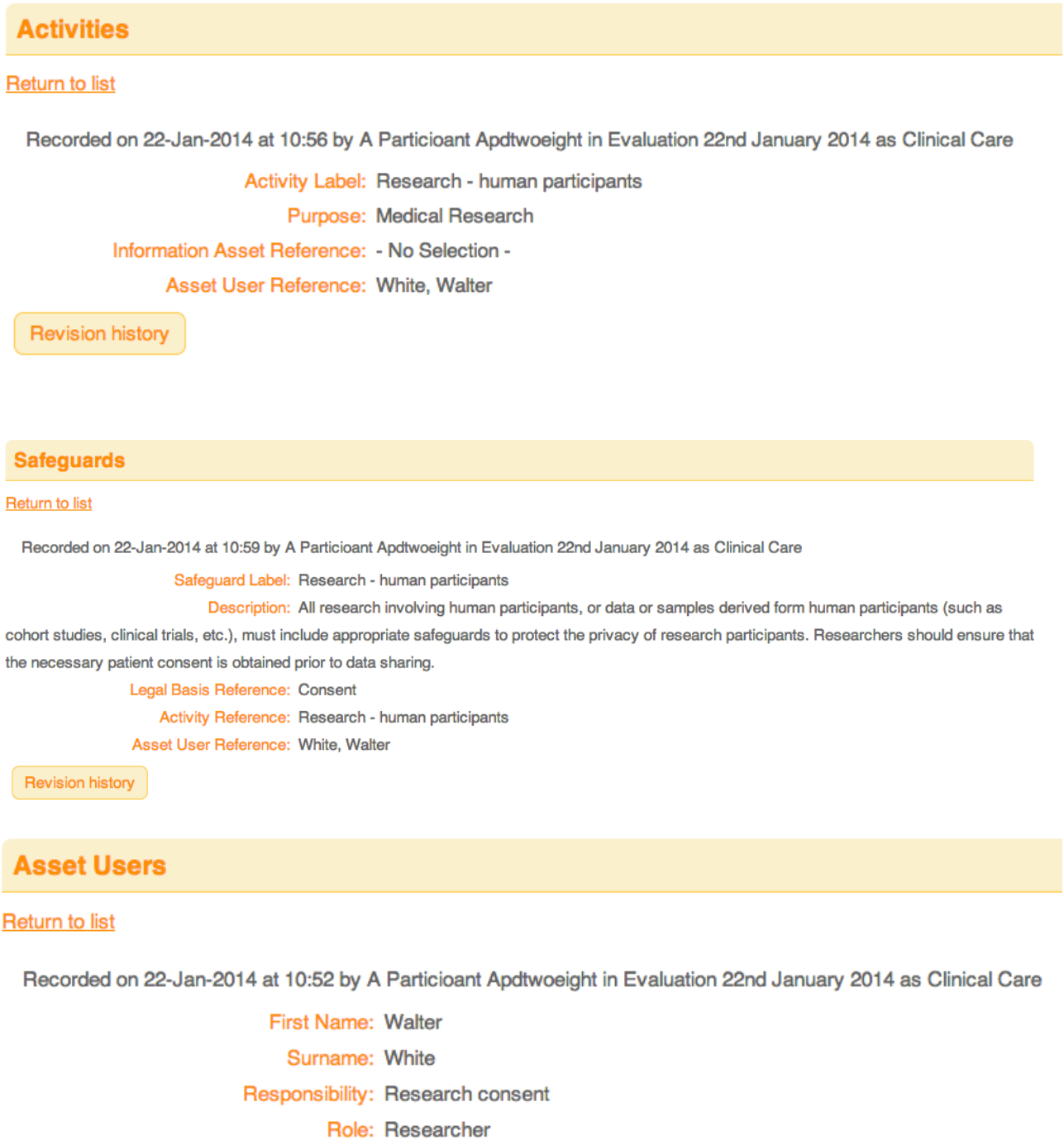
Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.



Response Review: the participant has correctly provided part of the Safeguard as expected. They have not specified a Control, however, and have not referred to the pre-prepared details.

Question 5: All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include

appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.



Activities

Return to list

Recorded on 22-Jan-2014 at 10:56 by A Particioant Apdtwoeight in Evaluation 22nd January 2014 as Clinical Care

Activity Label: Research - human participants
Purpose: Medical Research
Information Asset Reference: - No Selection -
Asset User Reference: White, Walter

Revision history

Safeguards

Return to list

Recorded on 22-Jan-2014 at 10:59 by A Particioant Apdtwoeight in Evaluation 22nd January 2014 as Clinical Care

Safeguard Label: Research - human participants
Description: All research involving human participants, or data or samples derived form human participants (such as cohort studies, clinical trials, etc.), must include appropriate safeguards to protect the privacy of research participants. Researchers should ensure that the necessary patient consent is obtained prior to data sharing.
Legal Basis Reference: Consent
Activity Reference: Research - human participants
Asset User Reference: White, Walter

Revision history

Asset Users

Return to list

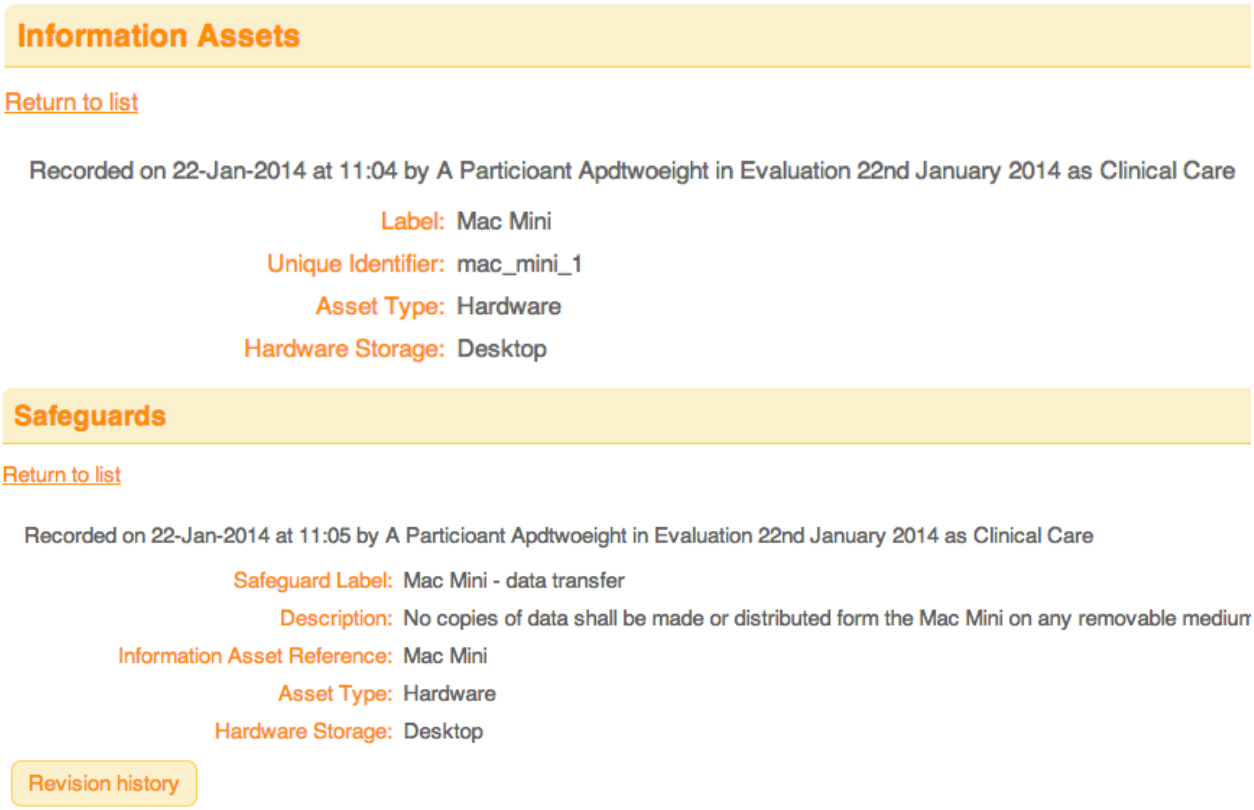
Recorded on 22-Jan-2014 at 10:52 by A Particioant Apdtwoeight in Evaluation 22nd January 2014 as Clinical Care

First Name: Walter
Surname: White
Responsibility: Research consent
Role: Researcher

Response Review: the participant has added an Asset User as expected, with appropriate details. They have also provided a Safeguard with appropriate details,

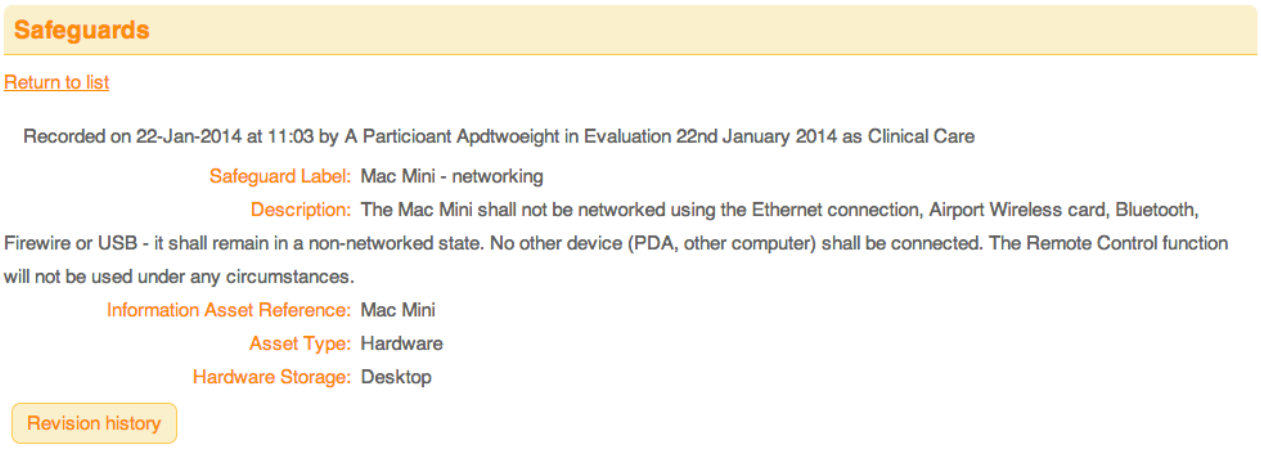
though have not included a Control. They have also supplied an Activity for Research, though this should have been provided in Question 4. It is nevertheless reasonable, with the exception of the No Selection for Information Asset reference.

Question 6: The Mac Mini shall not be networked using the Ethernet connection, Airport Wireless card, Bluetooth, Firewire or USB - it shall remain in a non-networked state. No other device (PDA, other computer) shall be connected. The Remote Control function will not be used under any circumstances.



Response Review: The participant has supplied an Information Asset with reasonable details and a Safeguard, without any Control. They have also not provided an Information Asset in the form of a PDA.

Question 7: No copies of data shall be made or distributed from the Mac Mini on any removable medium.



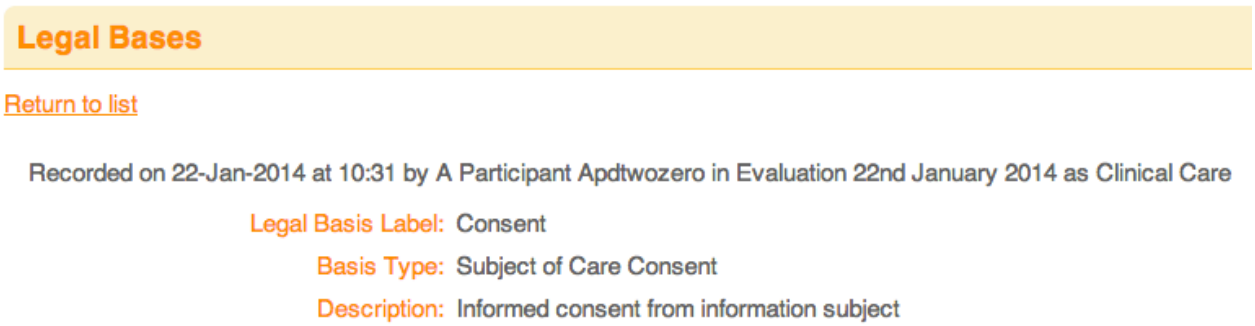
Response Review: The participant has supplied a Safeguard with appropriate details, though again no Controls. They should have included the CD ROM information Asset reference as well..

Participant APD_00020

Question 1: Please enter the information asset "CD ROM" into keibi, completing the details as you see fit.

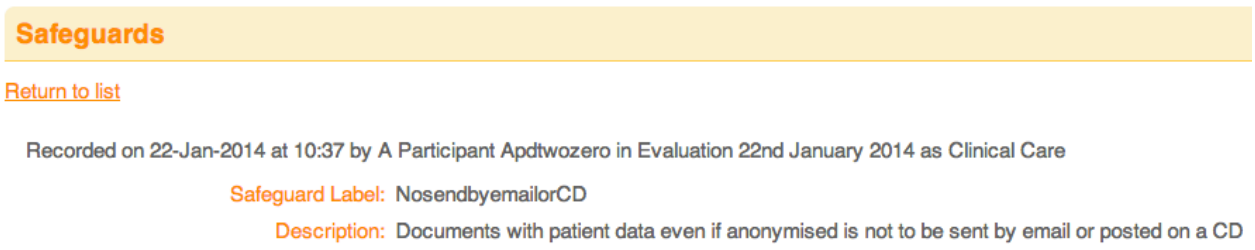
Response Review: The participant has specified the Information Asset details.

Question 2: Please enter the Legal Basis "Consent" into keibi, completing the details as you see fit.



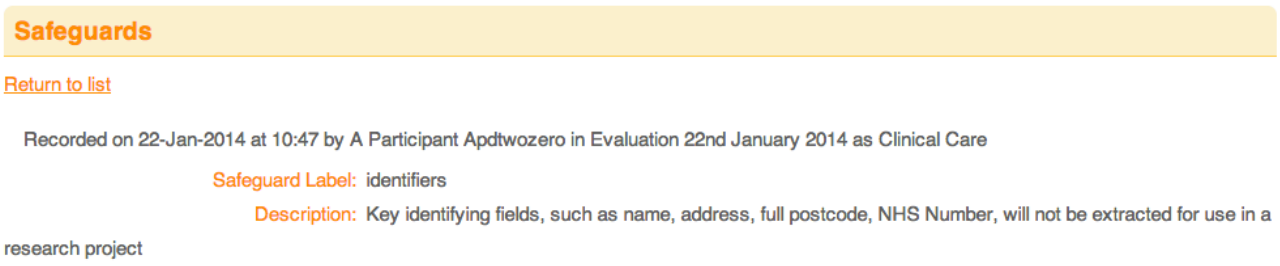
Response Review: The participant provided the correct details for this question.

Question 3: Documents with patient data even if anonymised is not to be sent by email or posted on a CD.



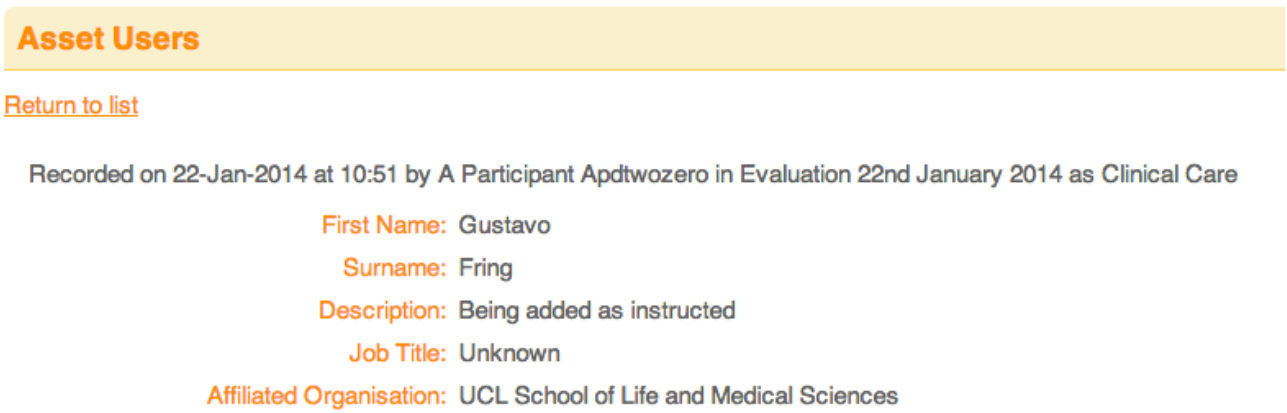
Response Review: The participant has specified a Safeguard with correct details, though there is no reference to the CD ROM Information Asset, and no Control. Additionally, the Activity of Sharing Information has also not been provided.

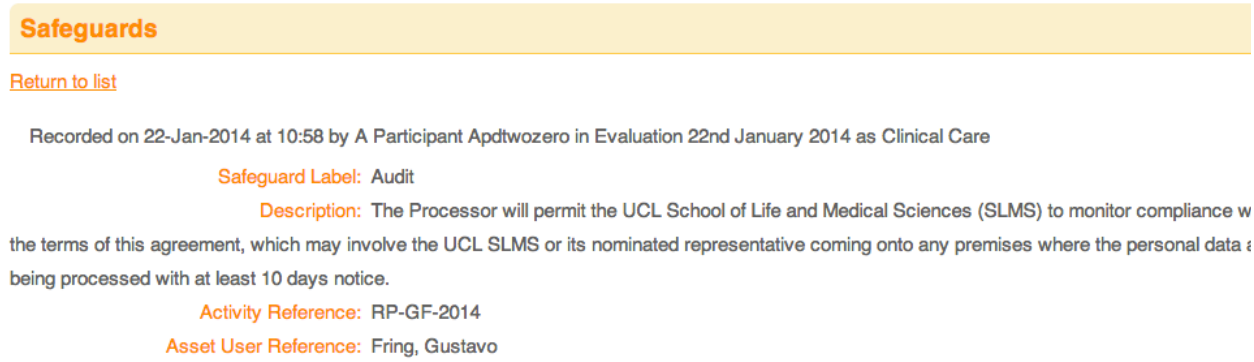
Question 4: Key identifying fields, such as name, address, full postcode, NHS Number, will not be extracted for use in a research project.



Response Review: The participant has specified a Safeguard with a Label and Description, but no Control and no reference to the identifying fields that were added. There is also no reference to the Activity of sharing information, which has not been provided, nor a reference to Research.

Question 5: Audit: The Processor will permit the UCL SLMS to monitor compliance with the terms of this agreement, which may involve the UCL SLMS or its nominated representative coming onto any premises where the personal data are being processed with at least 10 working days' notice.





Safeguards

Return to list

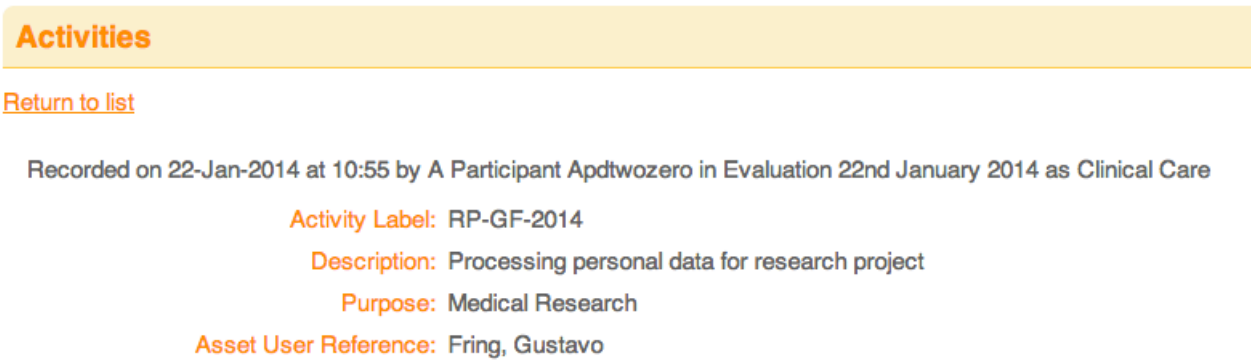
Recorded on 22-Jan-2014 at 10:58 by A Participant Apdtwozero in Evaluation 22nd January 2014 as Clinical Care

Safeguard Label: Audit

Description: The Processor will permit the UCL School of Life and Medical Sciences (SLMS) to monitor compliance w the terms of this agreement, which may involve the UCL SLMS or its nominated representative coming onto any premises where the personal data a being processed with at least 10 days notice.

Activity Reference: RP-GF-2014

Asset User Reference: Fring, Gustavo



Activities

Return to list

Recorded on 22-Jan-2014 at 10:55 by A Participant Apdtwozero in Evaluation 22nd January 2014 as Clinical Care

Activity Label: RP-GF-2014

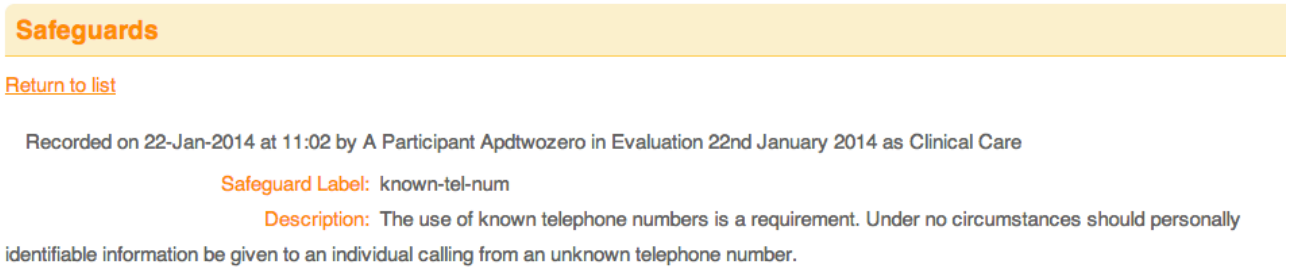
Description: Processing personal data for research project

Purpose: Medical Research

Asset User Reference: Fring, Gustavo

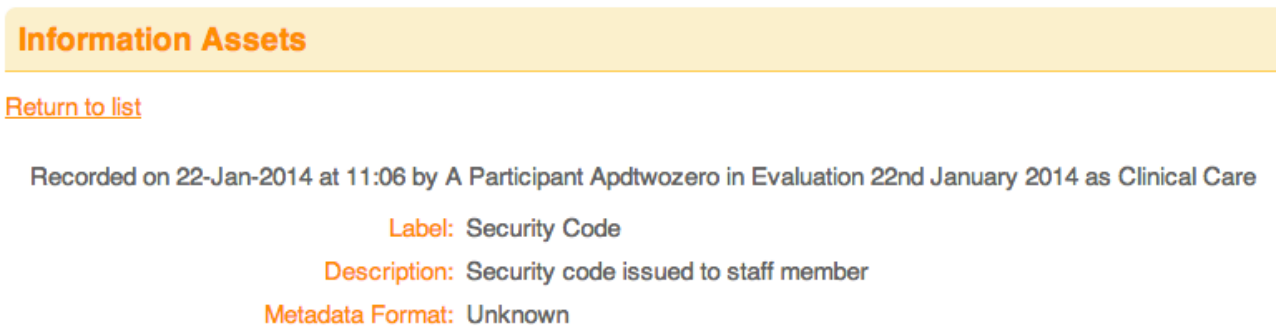
Response Review: The participant has added an Asset User, with reasonable details except for Description and Job Title - the Asset User is an Auditor. They have however unexpectedly entered the Activity of Processing personal data for research, and made this an audit able Activity as per the Supplied Safeguard. the details here are correct, with the exception of the Addition of Gustavo Fring as a researcher - they are an Auditor.

Question 6: The use of known telephone numbers is a requirement. Under no circumstances should personally identifiable information be given to an individual calling from an unknown telephone number.



Response Review: The participant has correctly specified a Safeguard, with only a label and Description. There is also no reference to an Information Asset for known telephone numbers, and they have not supplied an Information Asset for known telephone numbers.

Question 7: In the event that unauthorised access was gained through security codes or passes being compromised the Line Manager must immediately take appropriate action to prevent further unauthorised access. Security codes should be changed and any passes used for unauthorised access should be disabled.



Response Review: The participant has correctly specified a minimal Safeguard with reasonable details, but has only supplied a reference to one Information Asset and no Control. They did add the Information Assets for Security Codes and Passes, with reasonable details.

Experiment 2

Participant APD_00031

General Comments:

(For all the tasks in Experiment 2, all I did was look through the Safeguards. Other functionality not used.)

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

Would not agree. *keibi* Safeguard says as patient data (even anonymised) to be sent by email.

It is unclear from *keibi* whether I can ever share data.

It doesn't rule out sharing non-identifiable data, but nor does it say that is allowable. I would, however, consult my line manager.

Response Review: The expected response was that anonymised data could be released, but not shared via email or CD, based upon the stipulations authored by Participant APD_00020 in Questions three and four. The question was answered incorrectly with three responses: whilst it is true that patient data cannot be sent by email, the participant felt that it was unclear from *keibi* as to whether the anonymised data could be shared. This was clear in the Safeguards authored in question three. 00020

Question 2: You have received a telephone request for names and addresses of three participants in a research trial. How would you respond?

Response to Question 2:

Is this a known telephone number? If not, I say nothing (as per Safeguard#3). If yes… well, who is this person? Why do they want the data? Safeguard#5 says that

such data (being identifying) cannot be given out for research purposes. Policy otherwise unclear: doesn't rule out, doesn't rule in.

Response Review: This question expected that the participant would use the details added in questions five and six and they made six responses, two of which were correct and expected, the rest of which are unexpected and show that the participant used their own expertise and experience to provide an answer, as well as highlighting a lack of clarity on the matter in the policy.

Question 3: A team member from UCL CLMS has asked to come and inspect your compliance with a data sharing agreement you hold with them within the next week. How would you respond to their request?

Response to Question 3:

The participant chose option 4: *I would agree and offer dates ten or more days later.* This was the correct response.

Safeguard#4 says that they may inspect but only with as least 10 days notice.

However, I am unclear why such notice is required, so I'd ask my line manager whether an earlier date is OK if both parties (us & UCL SLMS) agree.

Response Review: Participant APD_00031 made the correct response and provided the correct justification for it. They also exceeded expectations by questioning the original policy and raising it with their line manager.

Question 4: You have received a report of an unauthorised access to an information asset. How do you respond to this?

Response to Question 4:

Panic. Fear. Sweating. Tachycardia. Contact my line manager about the situation, Panic some more. *keibi* says (first 2 Safeguards) that the line manager must take action - so I refer to the line manager.

keibi differentiates between codes and passes being compromised, so I would seek to determine codes &/or passes were involved, & supply my findings to the line manager.

Response Review: Participant APD_00031 made five responses, three of which were correct of which were unexpected based upon what was entered in *keibi*, where they seek to determine the problem areas and inform the line manager of these, surpassing expectations. The last two responses refer to emotional and physiological responses, which the investigator has opted to leave out of the analysis!

Participant APD_00028

General Comments:

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

According to the policy patient data (even if anonymised) is not to be sent by email; the data would have to be transferred differently. I do not see a problem with sending de-identified information over email, however to be sure (blood pressure could theoretically identify someone) and to protect valuable data I would do this by encrypted container.

Response Review: The expected response was that anonymised data could be released, but not shared via email or CD. The question was answered correctly with five responses, all of which were reasonable, three of which were unexpected and relied on the participant's expertise.

Question 2: You have received a telephone request for names and addresses of three participants in a research trial. How would you respond?

Response to Question 2:

I would first need to check that this is a known number. But in any case since this is a request for a research project I would have to decline. Presumably this is for recruitment, so we could arrange for their participation without needing to send their details to a third party.

Response Review: This question expected that the participant would use the details added in questions five and six and they made three responses, one of which was expected (checking the telephone number), the other two of which were not but were reasonable.

Question 3: A team member from UCL SLMS has asked to come and inspect your compliance with a data sharing agreement you hold with them within the next week. How would you respond to their request?

Response to Question 3:

The participant chose option 4: *I would agree and offer dates ten or more days later.* This was the correct response.

They have the right to do this, but I have the right of at least 10 days notice.

Response Review: Participant APD_00028 made the correct response and provided the correct justification for it.

Question 4: You have received a report of an unauthorised access to an information asset. How do you respond to this?

Response to Question 4:

If the person is staff I would have to notify his / her one manager who would need to take action to prevent further access. The system administrator would need to be notified, by the line manager, to disable any security passes and change any security codes.

Response Review: Participant APD_00028 made three responses, all of which were correct.00031

Participant APD_00020

General Comments:

Question 1: A colleague working with you on a research project has asked you to send them all the blood pressure readings that you have collected in your research data repository, but no identifying information. The colleague has asked you to send these in an email in a spreadsheet.

Response to Question 1:

By checking the Safeguards I can see that sending any patient data by email is forbidden. I would not agree to this specific request. However, there are no Safeguards preventing the sharing of this information. I would suggest an alternative transport mechanism i.e. encrypted file via UCL Dropbox.

Response Review: This question expected that the participant would use the details added in questions three and four. There are four responses: these are all correct and reasonable, one of which exceeds expectations by offering a reasonable alternative transport mechanism.

Question 2: You have been asked to transfer data from a Mac Mini holding a research data repository to a colleague using a USB key. How would you respond?

Response to Question 2:

The Safeguards appear contradictory. The first Safeguard states that no data will be copied from the Mac Mini. However, a further Safeguard suggests that data can be shared as long as all identifying information is removed. I am assuming that Safeguard 1 refers to full copies. In this case, I would create a fully anonymised data extract for transfer to my colleague.

Response Review: This question expected that the participant would use the details added in questions six and seven by Participant APD_00028 and refers to the expected Safeguards. There are 5 responses, three of which are correct and expected, two of which are not: by providing a reasonable alternative, the participant has exceeded expectations.

Question 3: A colleague has requested that you network the Mac Mini to perform a critical software update. How would you respond?

Response to Question 3:

Incorrect answer: "I would decline the request, citing the information security policy"

I would explore alternatives to networking. Can the upgrade be completed using removable media for instance? If there was no alternative I would consult the owners of the policy and carry out a risk assessment.

Response Review: This question expected that the participant would use the details added in question six by Participant APD_00015. The response is incorrect. Though the wrong option was selected, the justification for this is reasonable and makes reasonable points and is in keeping with option 4.

.

Question 4: A research colleague has asked for some patient data. They say that it can be anonymised and that they would be able to receive it via email, on a posted CD or they will be happy to collect it on an encrypted USB key. How do you respond?

Response to Question 4:

I would generate an anonymised data set and ask the requestor to pick it up on an encrypted USB key.

Response Review: This question expected that the participant would use the details added in question four, five, six and seven by Participant APD_00020. There are two responses, one of which (creating an anonymised data set) is reasonable and unexpected. It is not clear whether the transfer to an encrypted USB key would be done directly from the Mac Mini or directly downloaded, but a USB key should not be connected. The original excerpt is unclear, however, on whether encryption is appropriate.

Appendix 18. Results Gathered from Experiment One

Results from Experiment 1

Total number of Expected and Authored Compositions

This section provides an analysis of the total numbers of Compositions and Controls that were expected and compares them with the number actually added by participants during the first experiment, showing how the *keibi* and the Secutype Knowledge Model was used in practice. These results are split between the two styles of exercise sheet where appropriate. Chart 2 below provides the total number of Expected Compositions, and Chart 3 and Chart 4 the breakdown between sheets one and two respectively. In total across the two exercise sheets, the evaluations results expected that participants would author twelve Safeguards, thirteen Controls, eight Activities, eleven Information Assets and two Legal Bases and Asset Users in order to answer the questions in the first experiment.

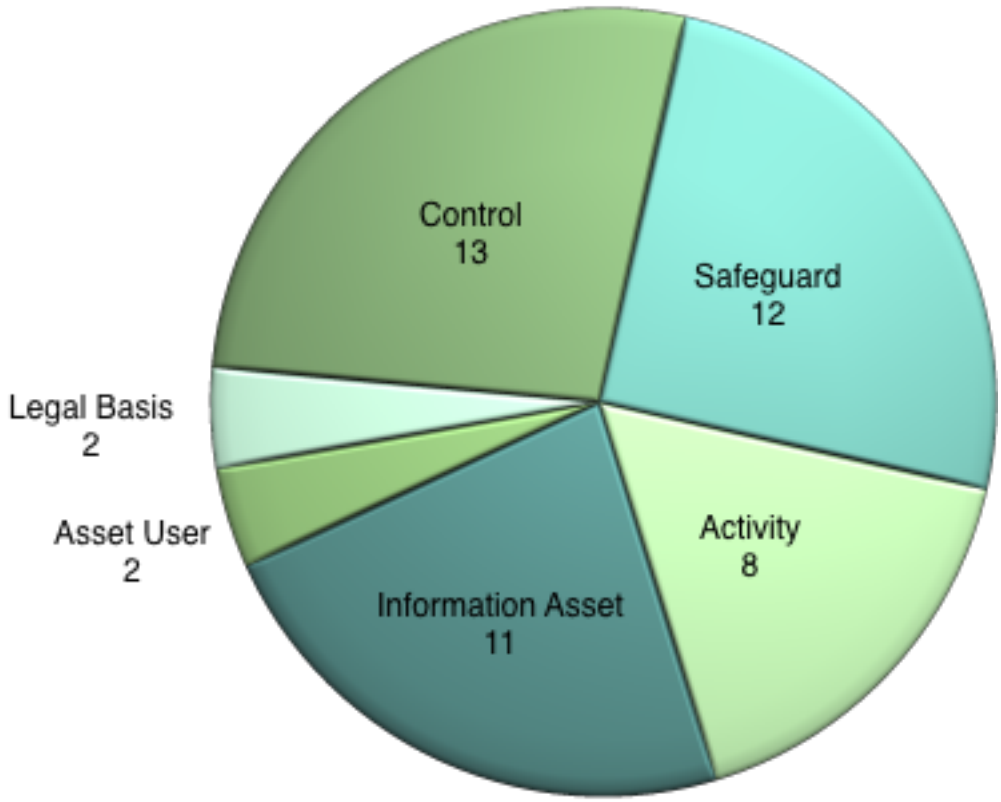


Chart 2: Breakdown of Expected Compositions for all exercise sheets

Across the two exercise sheets, six safeguards, four Activities, one Asset User and one Legal Basis were expected; Sheet 2 expected an additional Control and three additional Information Assets above those form Sheet 1, based upon the requirements of the excerpt used in the seventh question that had been provided.

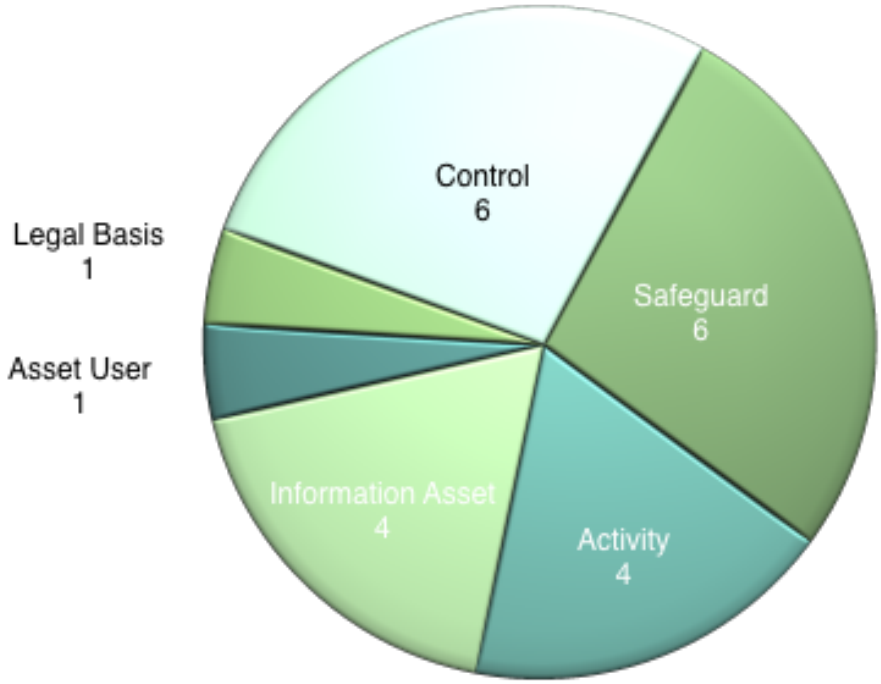


Chart 3: Breakdown of Expected Compositions, Exercise Sheet 1

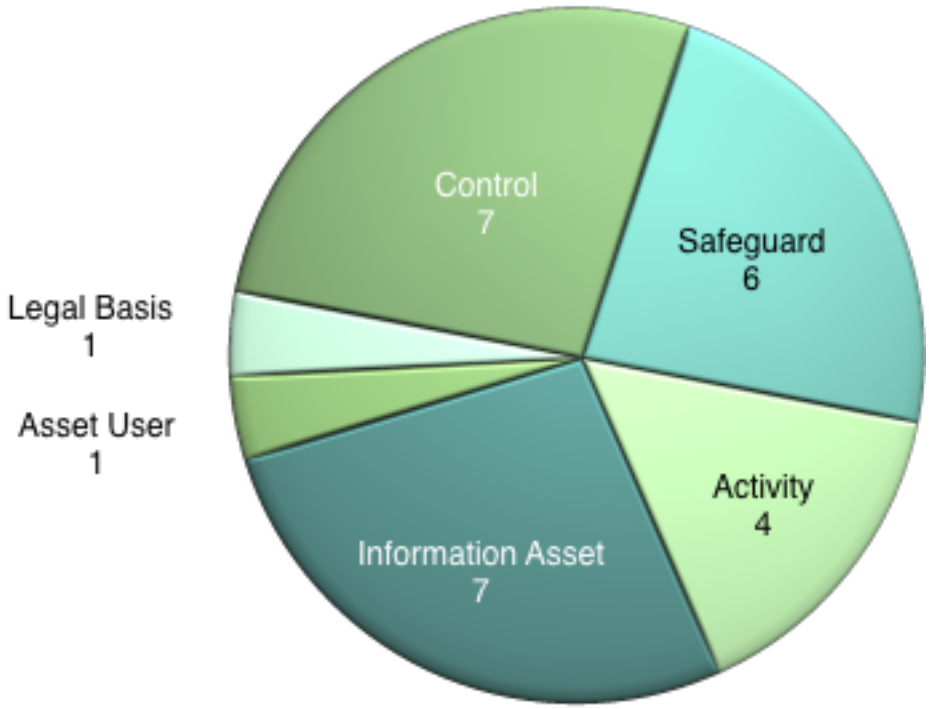
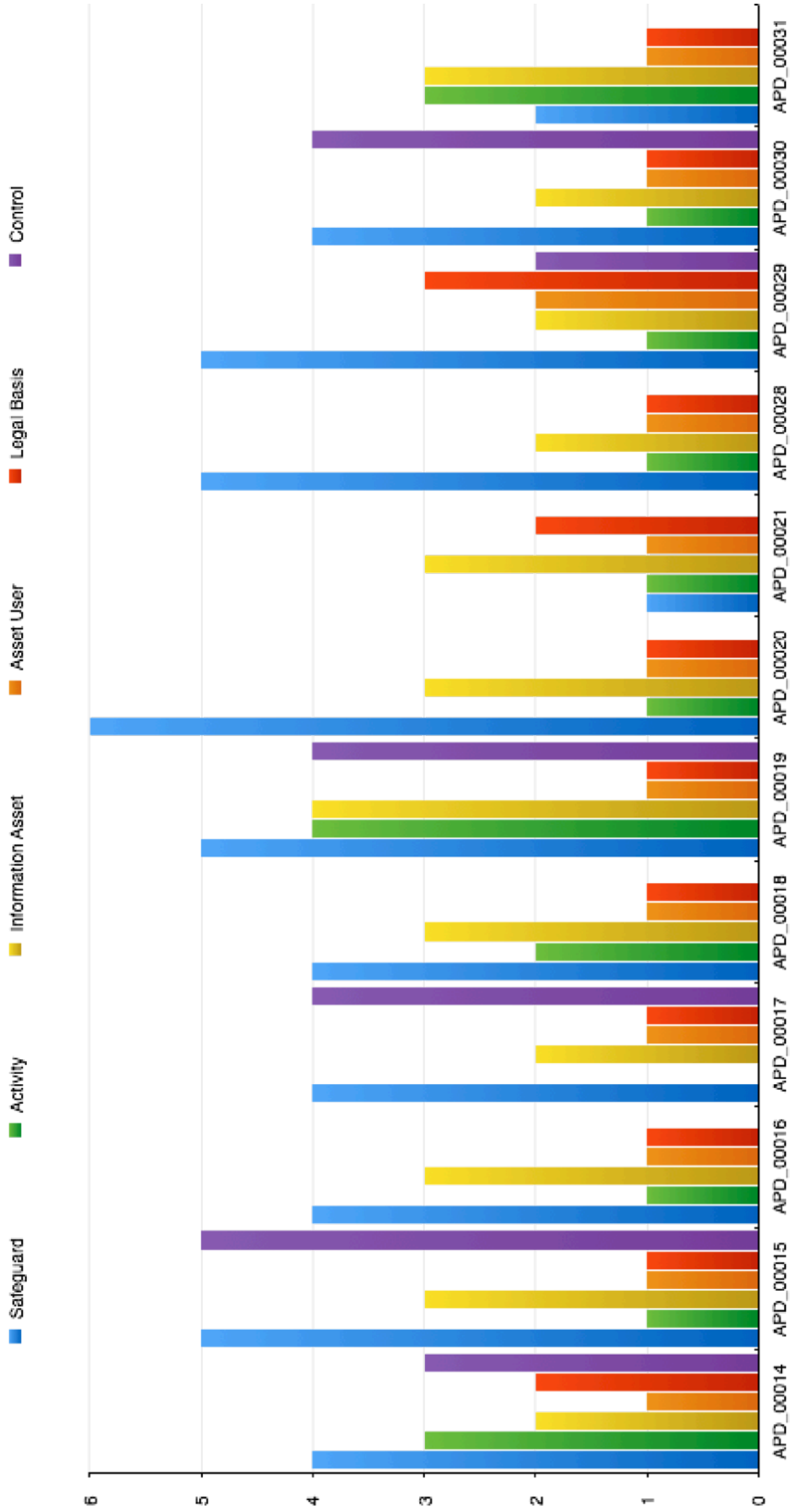


Chart 4: Breakdown of Expected Compositions, Exercise Sheet 2

Graph 10 below provides the total number of actual Compositions and Controls entered by participants. It is apparent from these results that participants deviated from the expected responses, where few participants added the expected number of Compositions or Controls, and generally added fewer than expected.



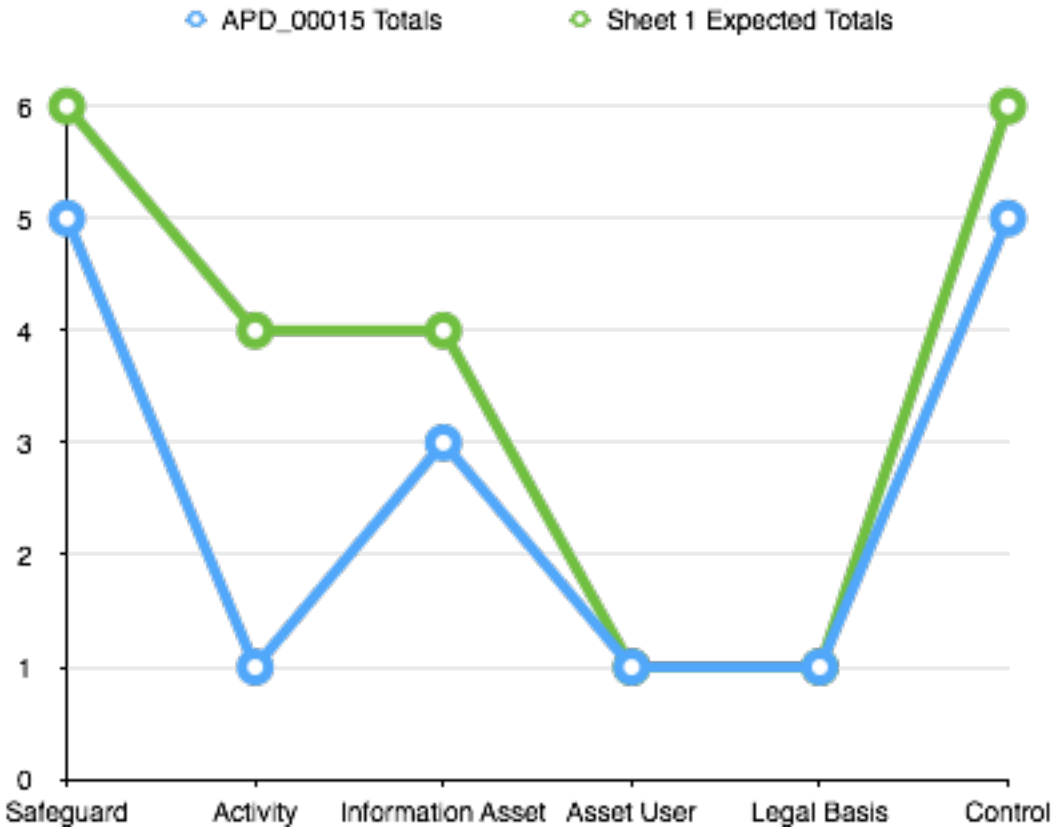
Graph 10: Total Number of Actual Compositions Entered by Participants in Experiment 1

Only one participant added the six expected Safeguards and one other participant added the expected number of four Activities. Only one participant added the expected number of Information Assets, whilst none added the expected number of Controls. Participants seemed to add the expected number of Asset Users and Legal Bases, though they were directed to do this in the question sheets.

The following two subsections contain twelve graphs, which break down the number of actual Compositions added for each participant and compares them with the expected number in each case. Each section provides the details for Sheets One and Two respectively. The next section further reviews the responses to assess understanding and correctness are assessed based upon the details that have been entered into each of the 13606 Classes that have been added.

Breakdown of Actual Composition Numbers Compared to Expected – Sheet 2

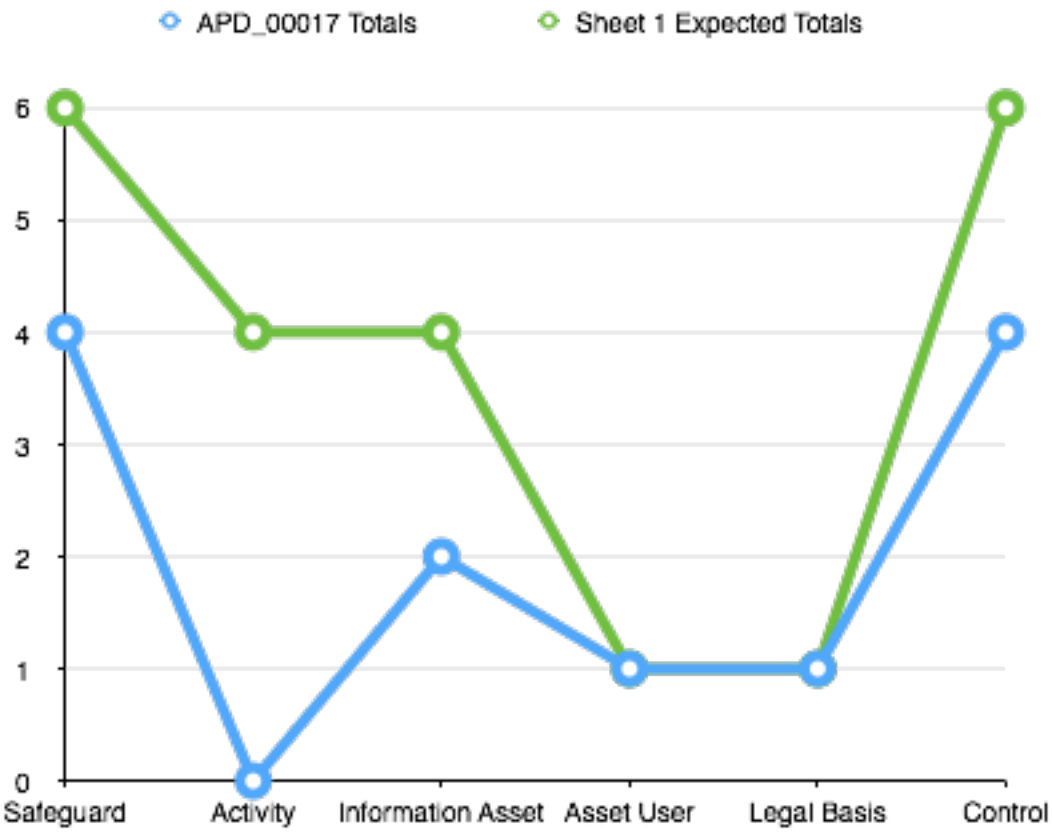
The total numbers of added Compositions and Controls are further analysed in the following subsection by reviewing the responses for each question from Sheet One, where the actual number of Compositions and Controls added are compared to the expected numbers for each participant.



Graph 11: Total number of Compositions and Controls added by APD_00015 compared with Expected Numbers

Graph 11 shows the number of Compositions and Controls added by Participant APD_00015 during the first experiment. It shows that they provided close to the expected number of Safeguards, Information Assets and Controls, with the expected number of Asset Users and Legal Bases (though they were directed to add these two Compositions in the question sheet). They had only one Activity, however.

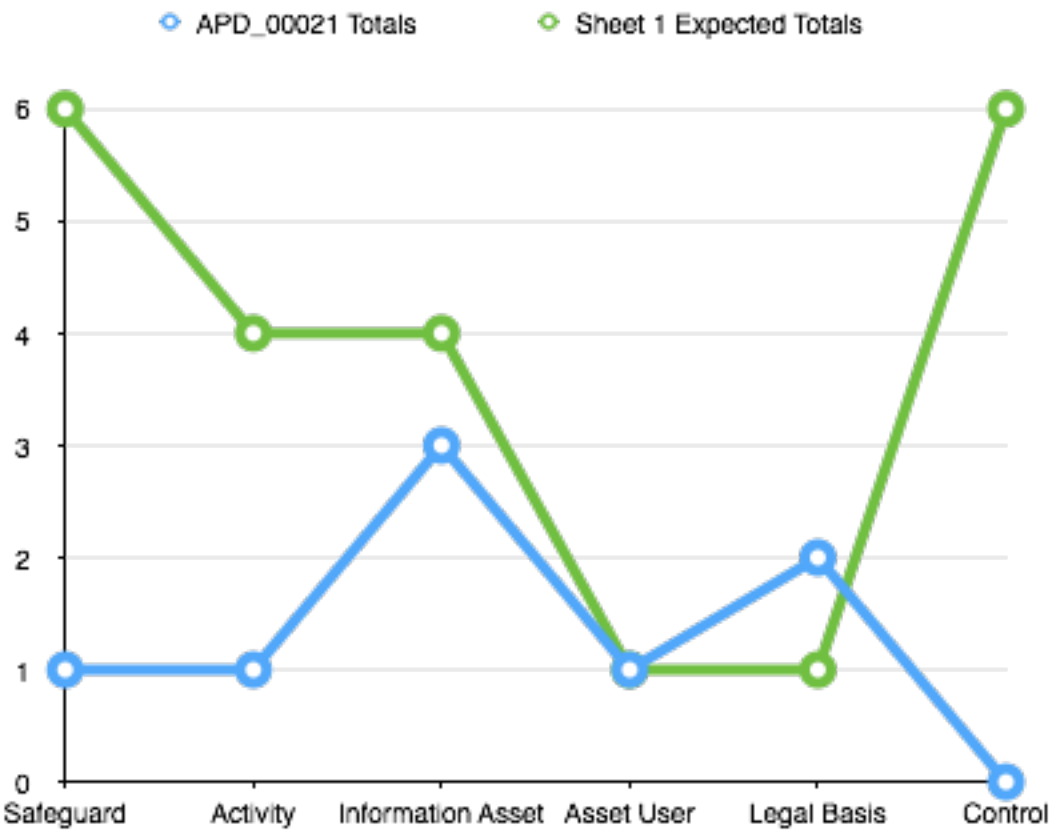
Graph 12 shows the comparison of total Compositions and Controls added compared with the expected number. There is some deviation from expectations here: there are no Activities in this case, four Safeguards instead of the expected six, two Information Assets instead of the expected four, and two fewer controls.



Graph 12: Total number of Compositions and Controls added by APD_00017 compared with Expected Numbers

Graph 13 provides the number of Compositions added by APD_00019. This participant was very close to matching the expected numbers of Compositions and Controls, with only one fewer Safeguard and Control than expected.

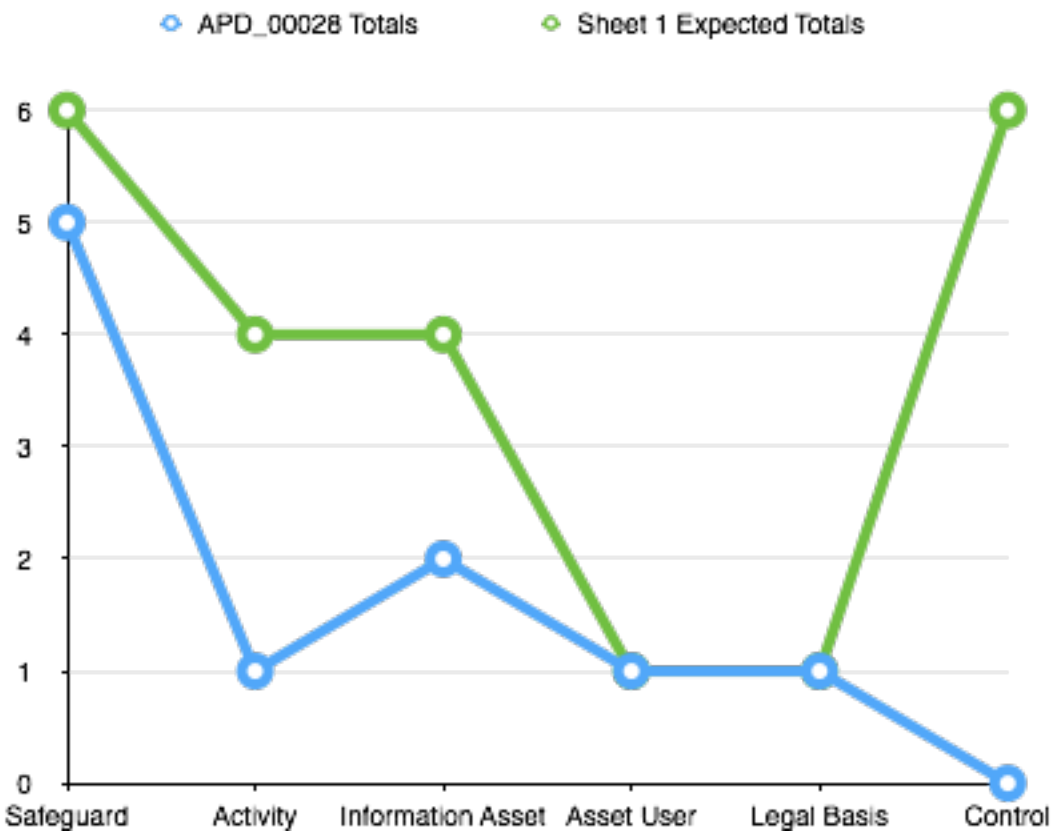
Graph 13: Total number of Compositions and Controls added by APD_00019 compared with Expected Numbers



Graph 14: Total number of Compositions and Controls added by APD_00021 compared with Expected Numbers

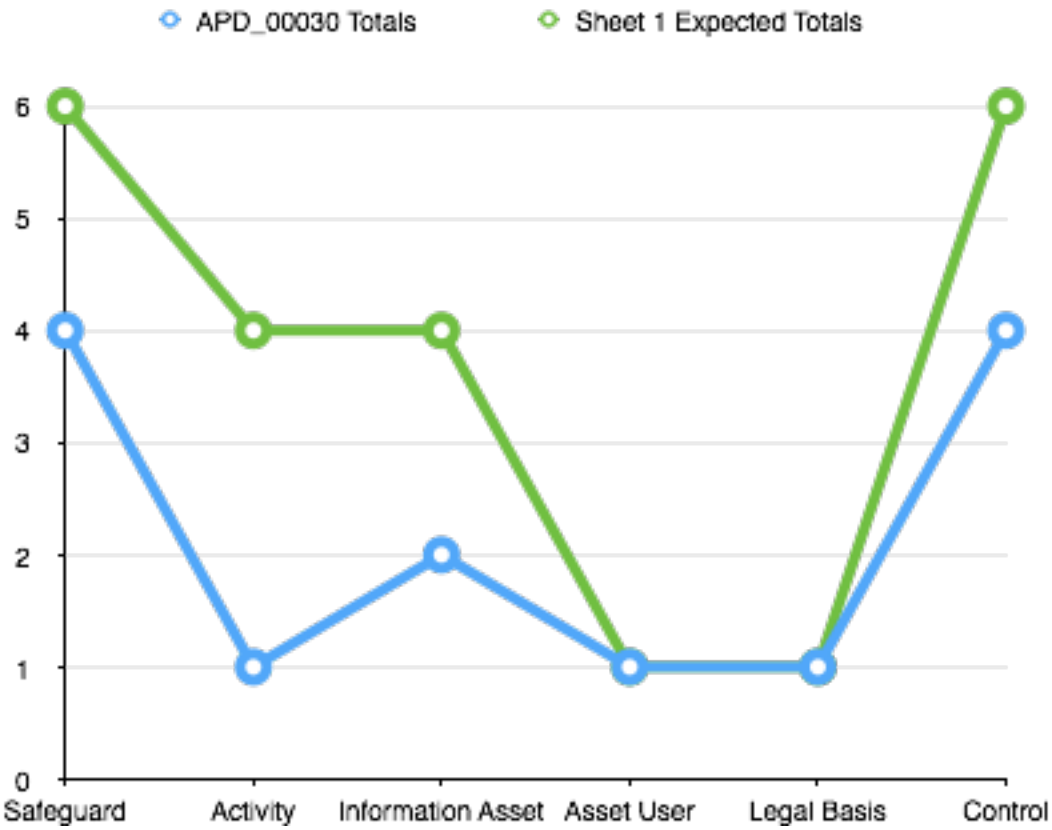
Graph 14 shows the comparison of actual Compositions and Controls added and expected outcomes for participant APD_00021. In this case, there is more deviation from expectation than the other cases, where there is only one Safeguard and no Controls. There is only one Activity and three Information Assets. There are however two Legal Bases where only one was expected. The analysis in the next section would provide some explanation for this.

Graph 15 provides the actual numbers added by participant APD_00028, comparing them with the expected outcomes. There is some deviation here from expectations, with only one Activity added, two Information Assets and no Controls. One Asset User and Legal Basis was added as expected.



Graph 15: Total number of Compositions and Controls added by APD_00028 compared with Expected Numbers

Graph 16 shows the comparison of actual numbers of Compositions and Controls added against the expected number for participant APD_00030. There is Again only one Activity and two Information Assets, one Asset User and Legal Basis, though there are this time four Controls, which is closer to expectations.



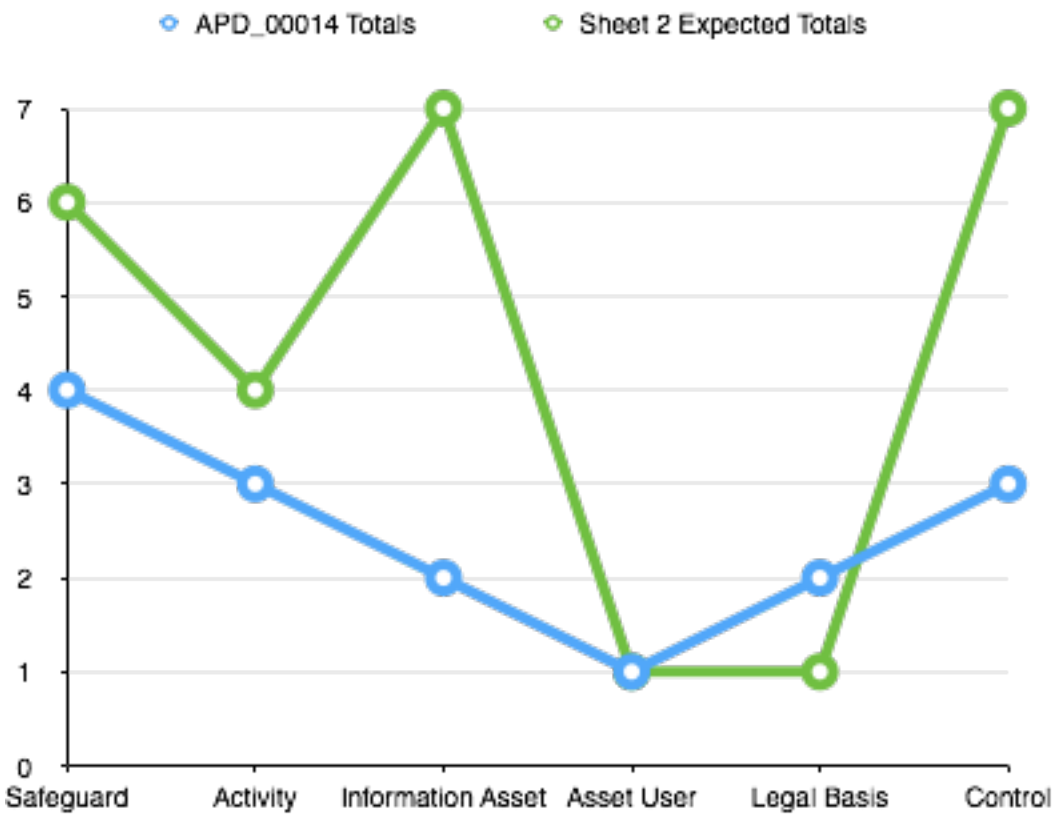
Graph 16: Total number of Compositions and Controls added by APD_00030 compared with Expected Numbers

Graph 17: Total number of Compositions and Controls added by APD_00031 compared with Expected Numbers

Graph 17 shows the actual number added by participant APD_00031, comparing these results with the expected number. A similar pattern has emerged with some of the other results, where significantly fewer Safeguards have been added than expected, and no Controls. There were three Activities and Information Assets in this case, however.

Breakdown of Actual Composition Numbers Compared to Expected – Sheet 2

The total numbers of added Compositions and Controls are further analysed in the following subsection by reviewing the responses for each question from Sheet Two, where the actual number of Compositions and Controls added are compared to the expected numbers for each participant.



Graph 18: Total number of Compositions and Controls added by APD_00014 compared with Expected Numbers

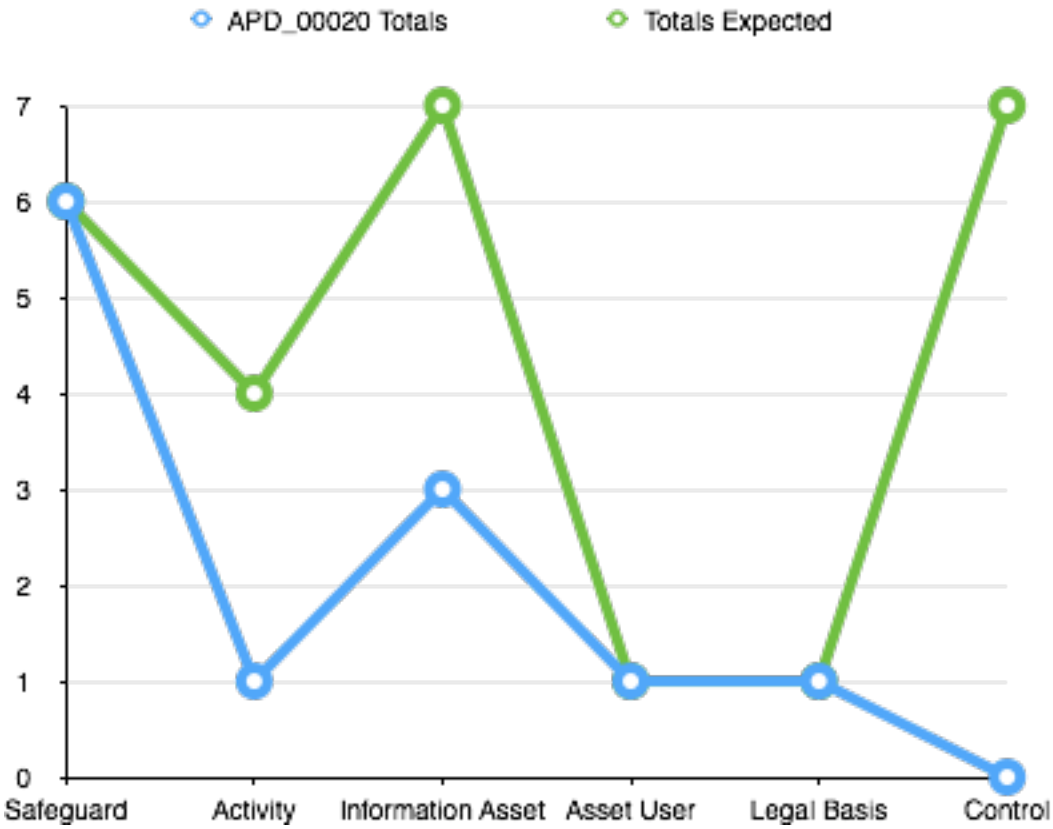
Graph 18 shows the total number of Compositions added by participant APD_00014. It is clear that they added fewer Safeguards (four), Activities (three) Information Assets (two) and Controls (three) than expected, though added an additional Legal Basis.

Graph 19 below shows the total number of actual Compositions authored compared with expected outputs for participant APD_00016. This shows a similar trend of adding fewer Compositions than expected with APD_00014, though in this case no Controls were added at all and the expected number of Legal Bases and Asset Users, as directed by the question sheet.

Graph 19: Total number of Compositions and Controls added by APD_00016 compared with Expected Numbers

Graph 20: Total number of Compositions and Controls added by APD_00018 compared with Expected Numbers

Graph 20 provides the comparison of actual Compositions and Controls added with the expected number for Participant APD_00018. There is a similar pattern here with participant APD_00016, with fewer Compositions being entered than expected, and no Controls. It should be noted that Participants APD_00016 and APD 00018 attended on different evaluation days.



Graph 21: Total number of Compositions and Controls added by APD_00029 compared with Expected Numbers

Graph 21 provides the numbers of Compositions added by Participant APD_00020 compared with the expected outcomes. This is a case where the participant added the expected number of Safeguards, though a familiar pattern has emerged in the case of the other Compositions, where fewer Activities and Information Assets were added than expected, and no Controls were added at all.

Graph 22 below shows the comparison for Participant APD_00029. In this case, the pattern seems to deviate form the others, where we have again fewer Safeguards, Activities and Information Assets than expected, though two Controls were added and a higher number of Asset Users and Legal Bases were added than

expected. The analysis of each response in the next section would reveal what participant APD_00029 had entered and whether they had understood the task they were completing.

Graph 22: Total number of Compositions and Controls added by APD_00029 compared with Expected Numbers

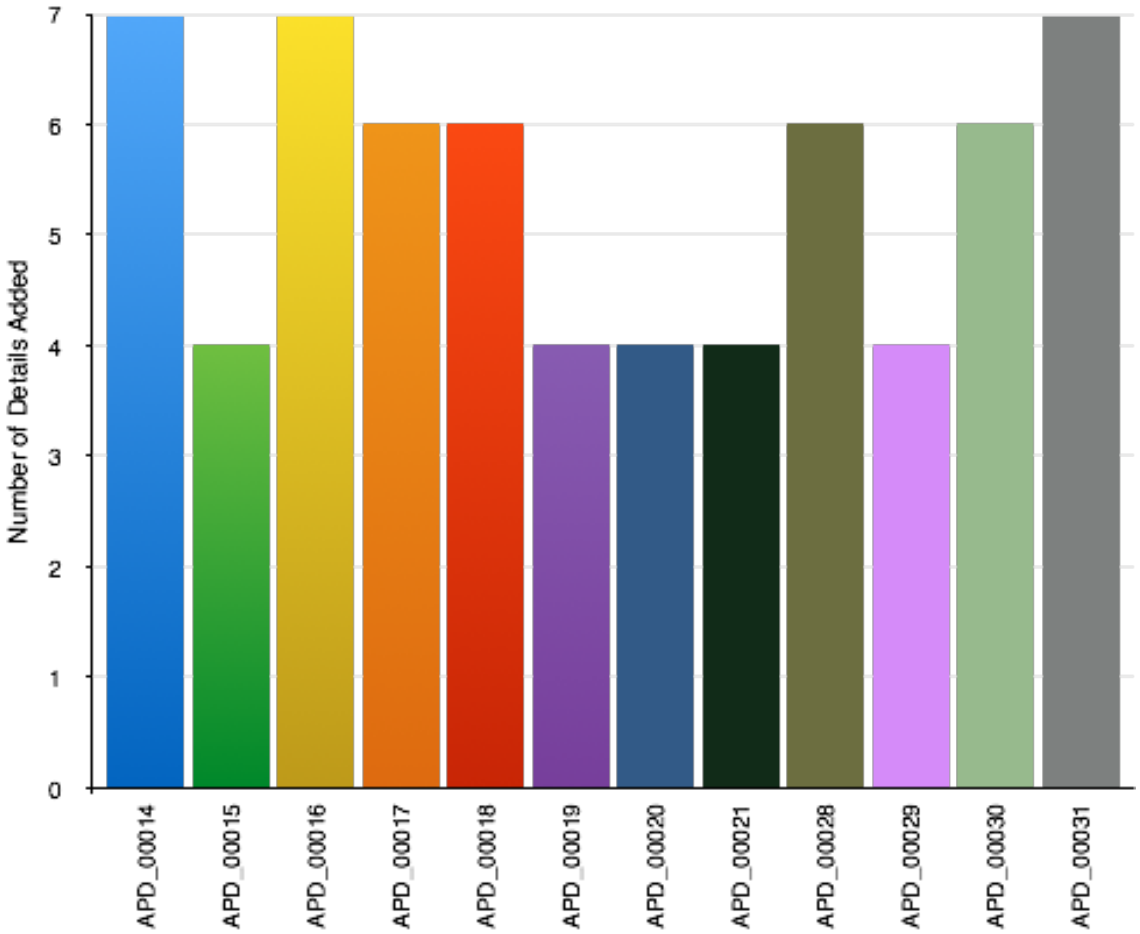
Comparison Across Participants

It is clear from the deviation between the expected numbers of Compositions and Controls and those participants opted to actually add that it was difficult to predict how participants would respond. Whether their responses demonstrated a lack of understanding, errors in authoring or an unpredicted approach to using *keibi* is reviewed in the next section. There are nevertheless some apparent trends illustrated across the participant results. There is a trend to under-specify Activities contrary to expectations, as well as Controls. There were more cases where there were fewer Compositions and Controls than expected, however expectations were slightly exceeded in the case of Asset Users and Legal Bases. The profiles of Composition and Control numbers do not suggest a consistent

pattern across the participant responses. Whilst this does suggest that further guidance and training might be appropriate, this could be contra-indicated in the results for assessing understanding and correctness of what was authored, as described in the next section.

Question 1 - CD ROM Information Asset

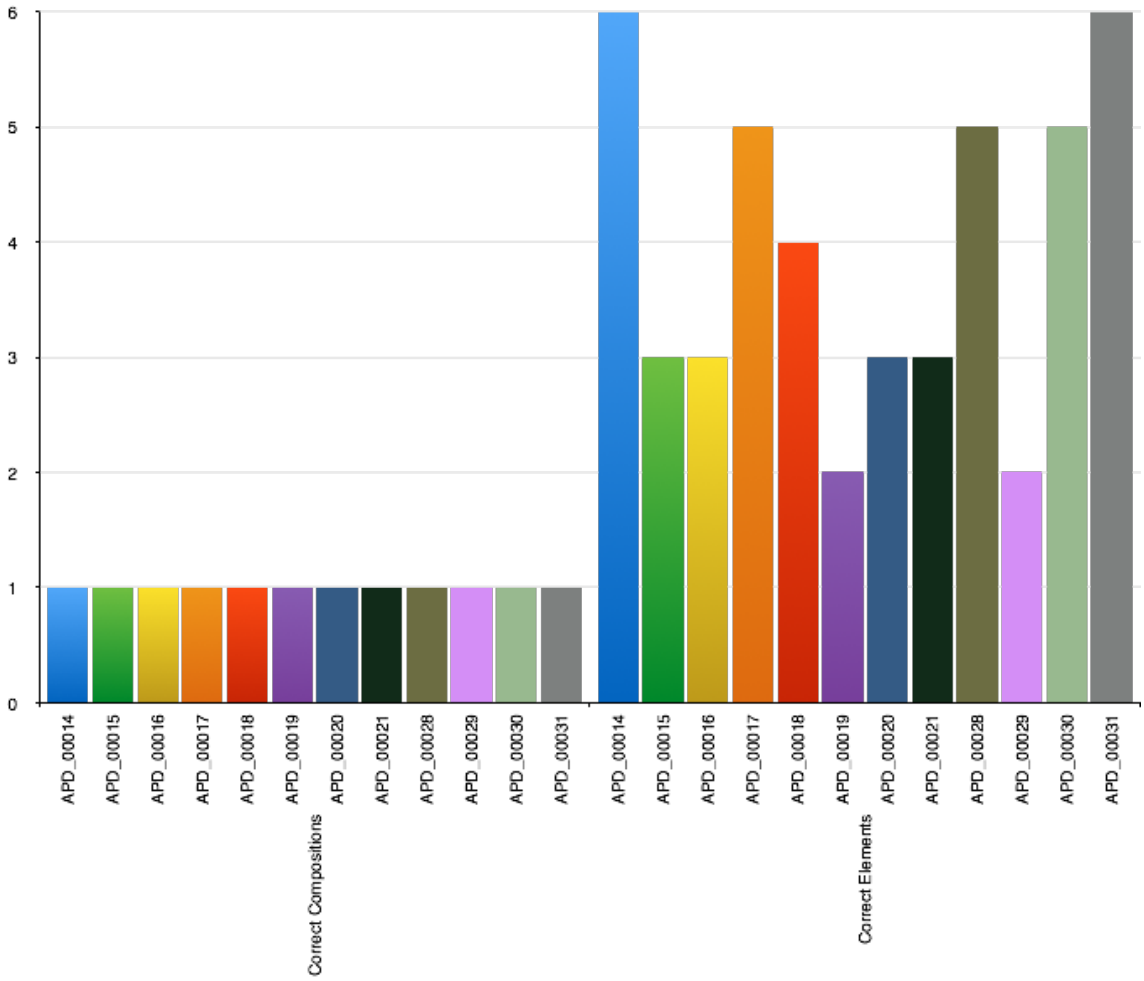
This question was designed to be a warm up for the rest of the experiment and Graph 23 shows that all participants understood this task and were able to add the details. Graph xxx shows that between four and seven Classes were added in total.



Graph 23: Number of Details added for Question 1

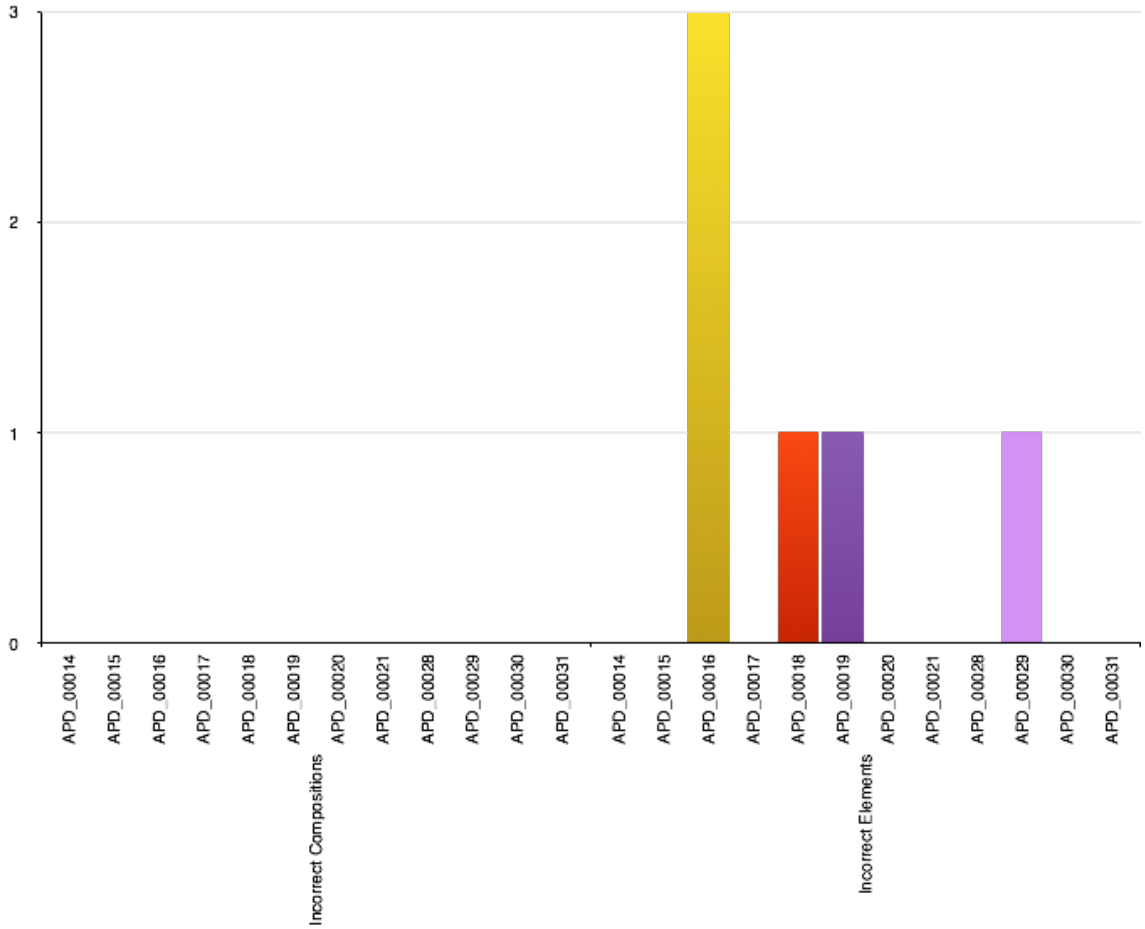
Graph 24 provides a breakdown of those Classes that were correctly added, with one correct Composition per participant and between two and six elements added. Graph 24 shows those Classes that were incorrectly added, with between one and three Elements added incorrectly across four participants. There were no

omissions in this question, and participants understood what they were expected to do.

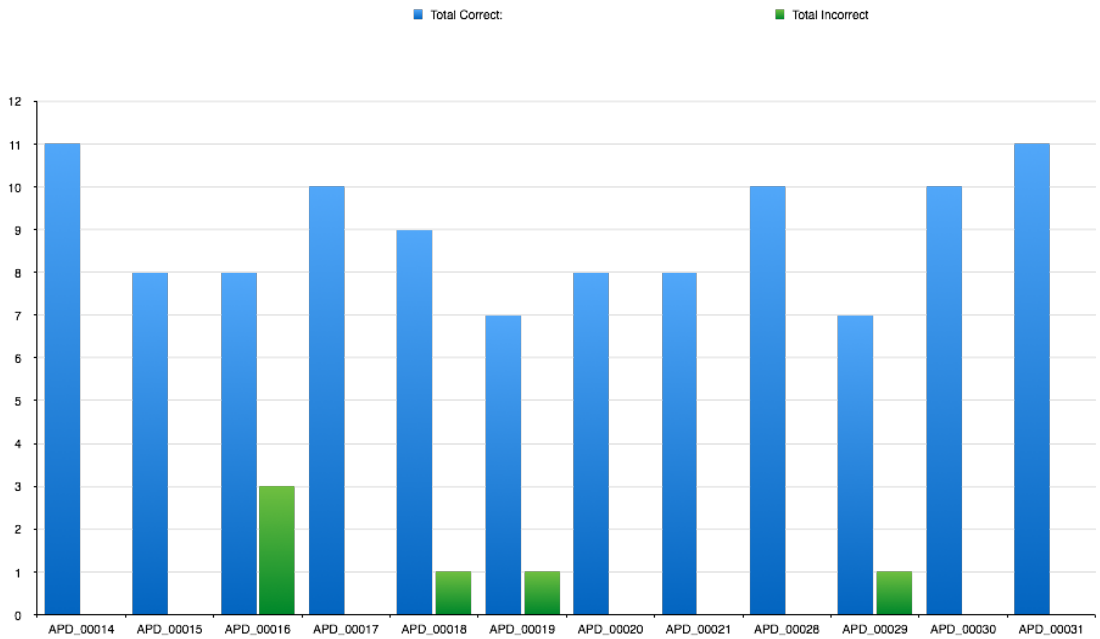


Graph 24: Number of correct 13606 Classes added in Question 1

Graph 26 shows the total correct and incorrect 13606 Class addition scores for question one across all participants. The maximum score was eleven and the minimum was seven. The maximum incorrect score was three and minimum was one. There were no omissions when reconciled with the expected responses, and it is clear that participants understood that they needed to add a single CD ROM information asset, as well as how to do that with the requisite Composition and Elements.



Graph 25: Number of incorrect 13606 Classes added in Question 1



Graph 26: Total Scores correct and incorrect classes added in Question 1

Question 2 - Consent Legal Basis

This question focused on the legal basis of consent and was intended as a warm up exercise. Chart 5 shows that there was some misunderstanding and exceeding expectations in this question, where there were ten examples of understanding and two of participant error misunderstanding. Graph 27 shows a consistent number of added 13606 Classes, with the majority of participants adding a total of four, and one adding three. In two cases, no Classes were added at all, which explains the misunderstanding as a result of participant error indicated in Chart 5: Totals for Understanding, Misunderstanding and Exceeding Expectations in Question 2.

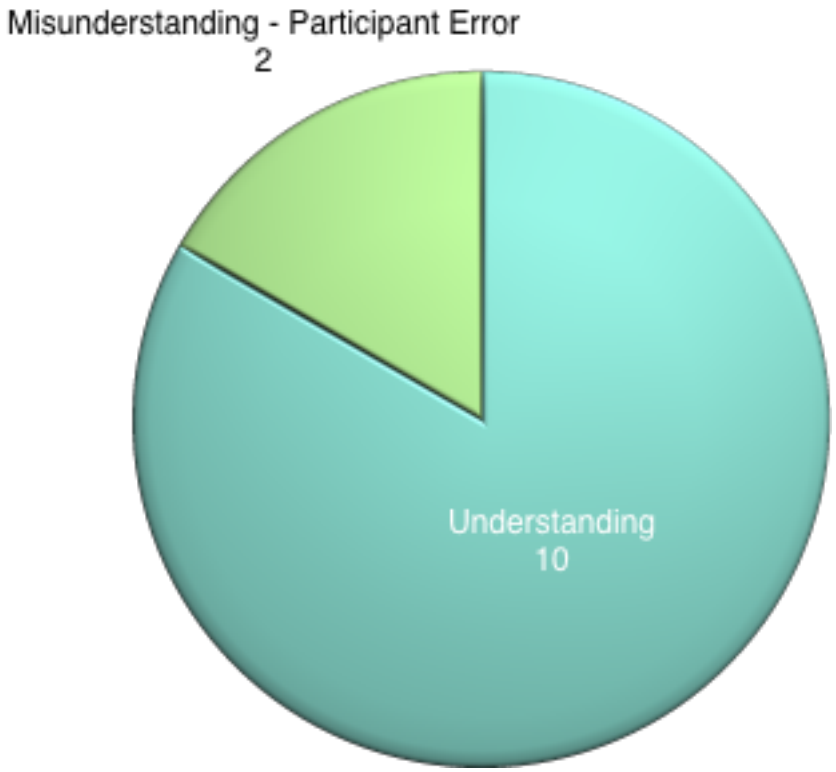
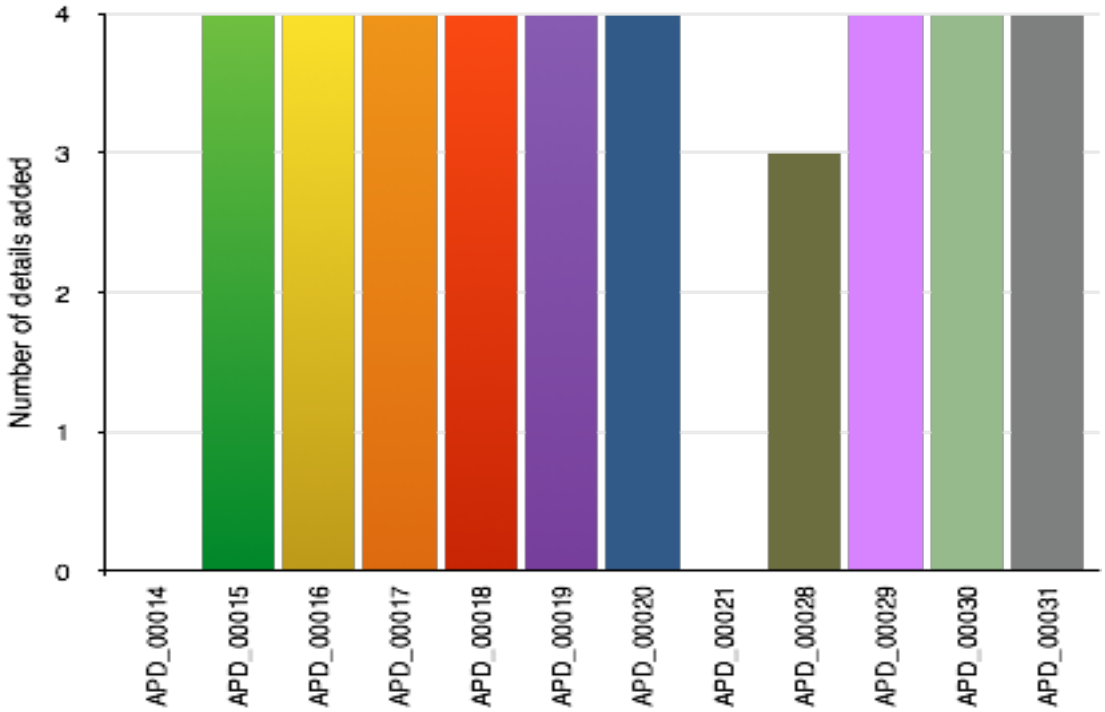


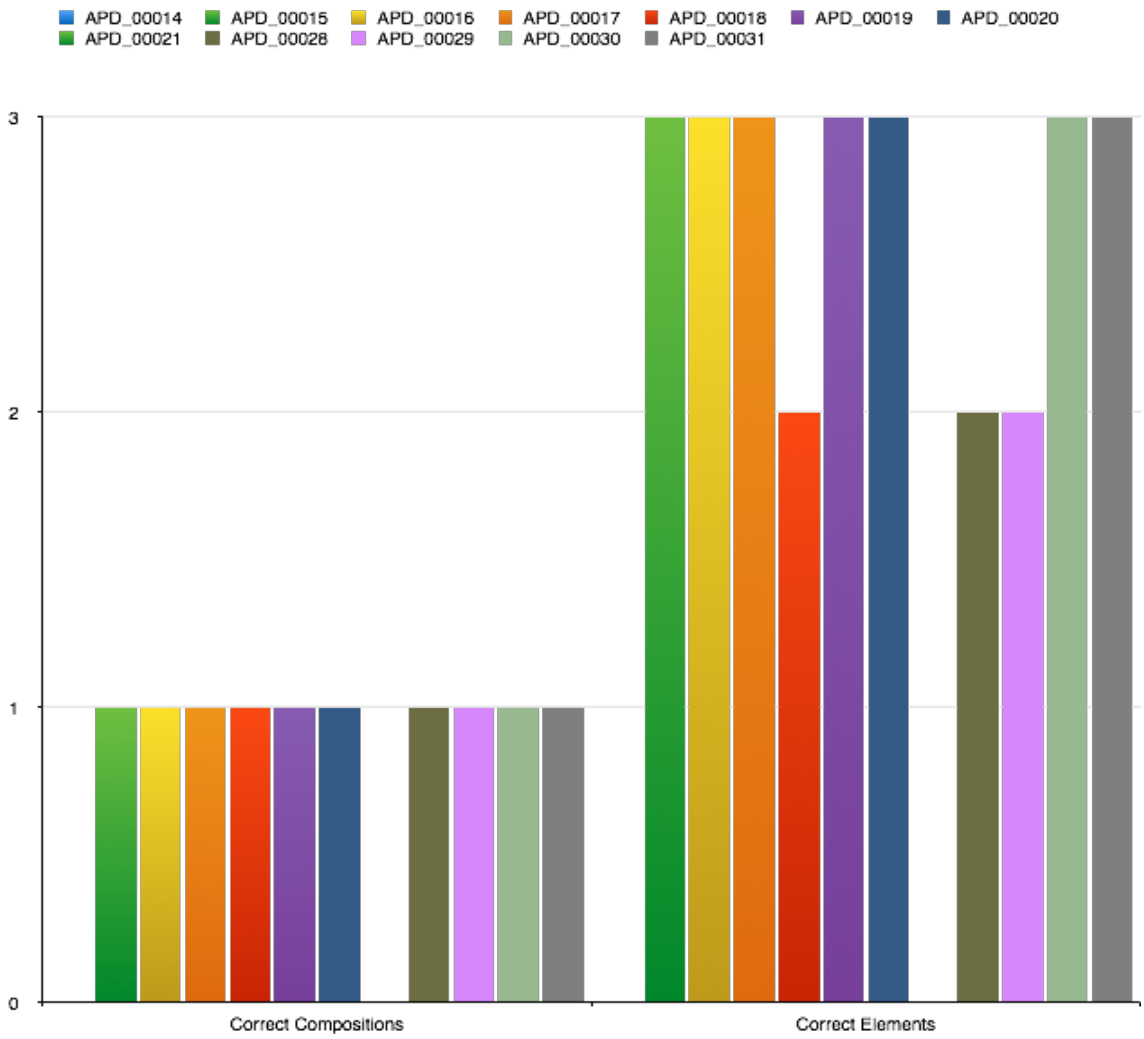
Chart 5: Totals for Understanding, Misunderstanding and Exceeding Expectations in Question 2

Graph 28 shows the number of correct 13606 classes added for question 2, where across the participants one Composition was added representing the legal basis, and between two and three elements were added. Graph 29 shows the number of incorrect Elements that were added, consistent with cases where two

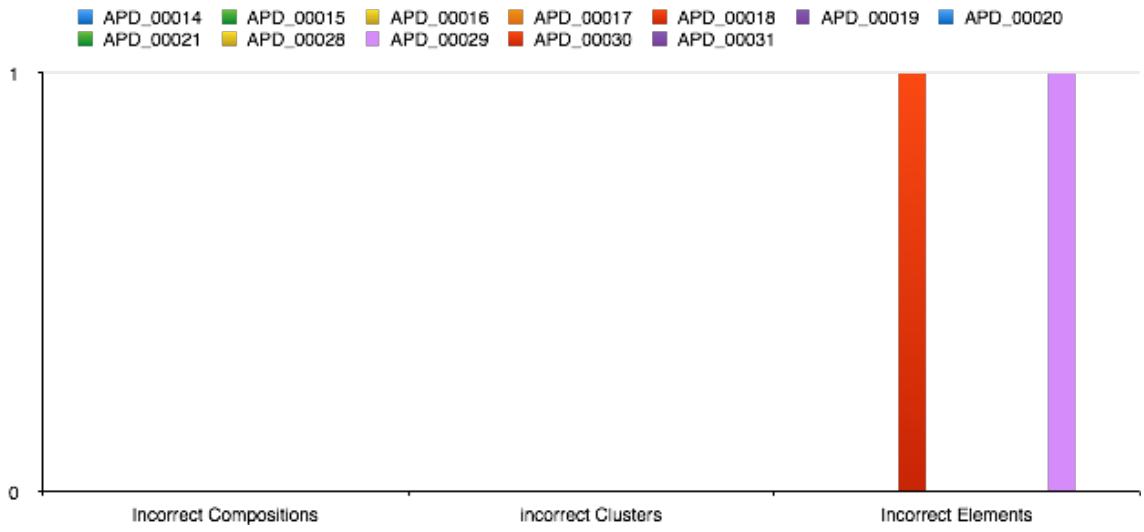
correct elements were added instead of three. The error from participant APD_00018 was because of an assumption that the consent related to the taking of clinical measures only, and APD_00029 only entered "dgdf" in the description field. Graph 30 shows the two cases where participants APD_00014 and APD_00021failed to enter a Consent Legal Basis.



Graph 27: Number of Details added in Question 2



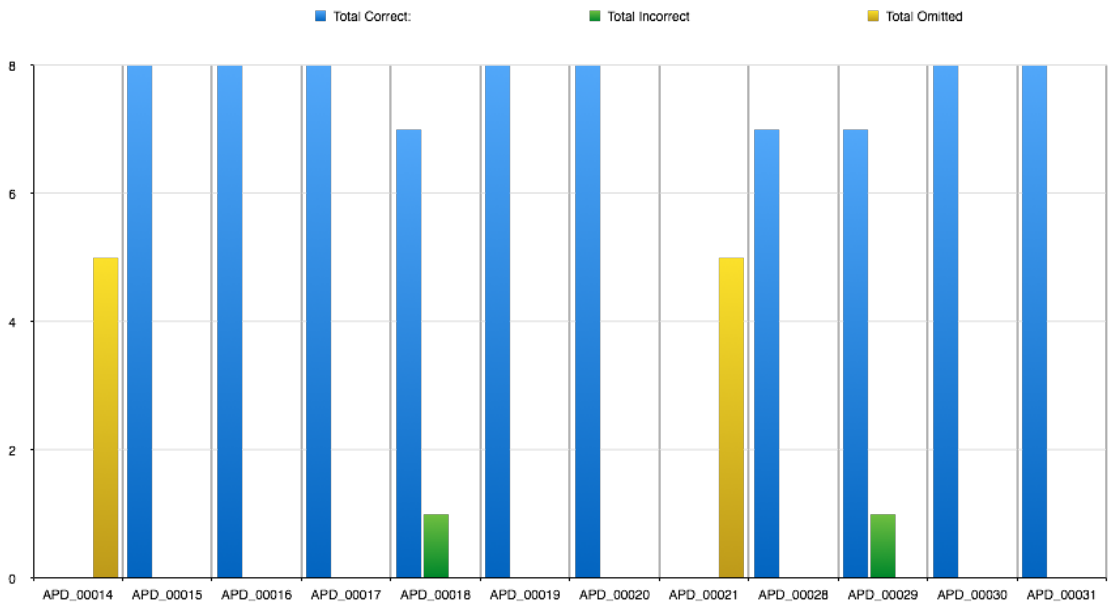
Graph 28: Number of correct 13606 classes added in question 2.



Graph 29: Number of incorrect 13606 classes added

Graph 30: Number of omitted 13606 classes in question 2

Graph 31 provides the total scores for correct, incorrect and omitted 13606 Classes, where the highest score was 8 and the lowest was zero. There were two cases of omitted Compositions and of incorrect Elements. The results showed that there was some confusion over adding Consent as a legal basis, indicating a possible uncertainty about terminology used to describe consent and translating that into the Legal Basis model that *keibi* provided.



Graph 31: Total Scores for Classes added in Question 2

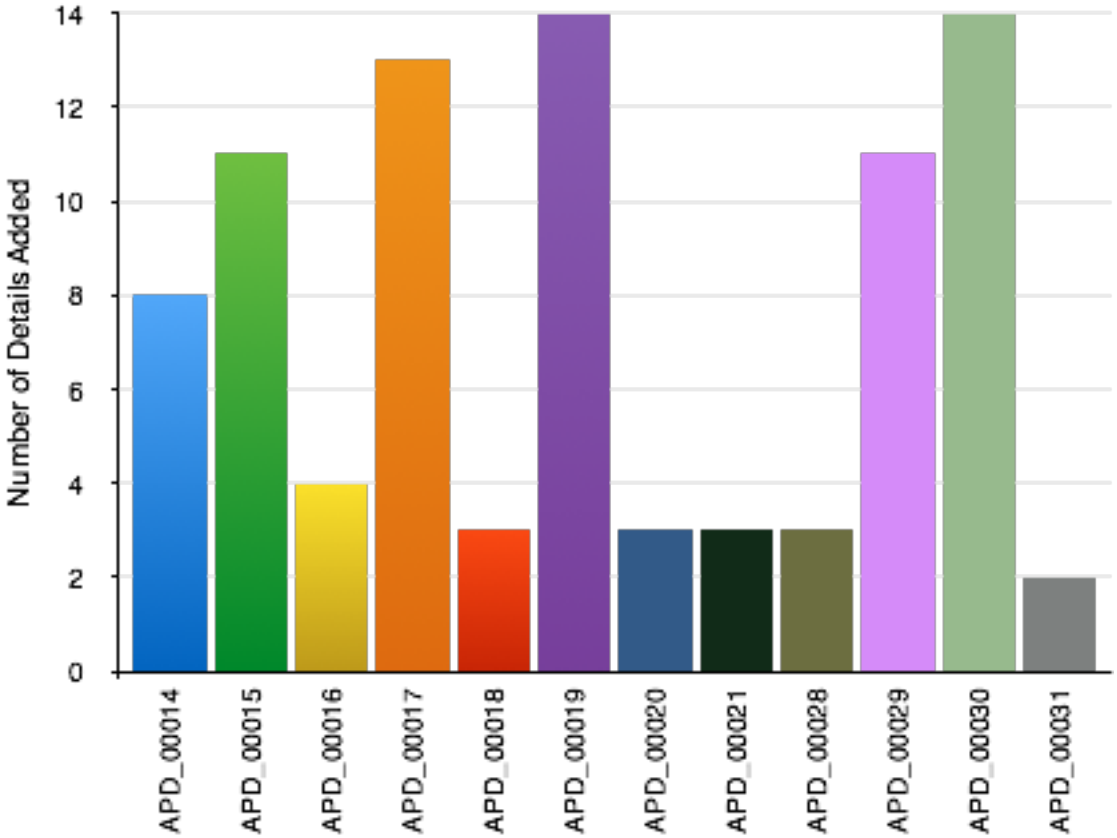
Question 3 - CD ROM Safeguard and Activity

This question showed more variation in understanding and the number of details added. Chart 6: Measure of understanding for each participant shows that there

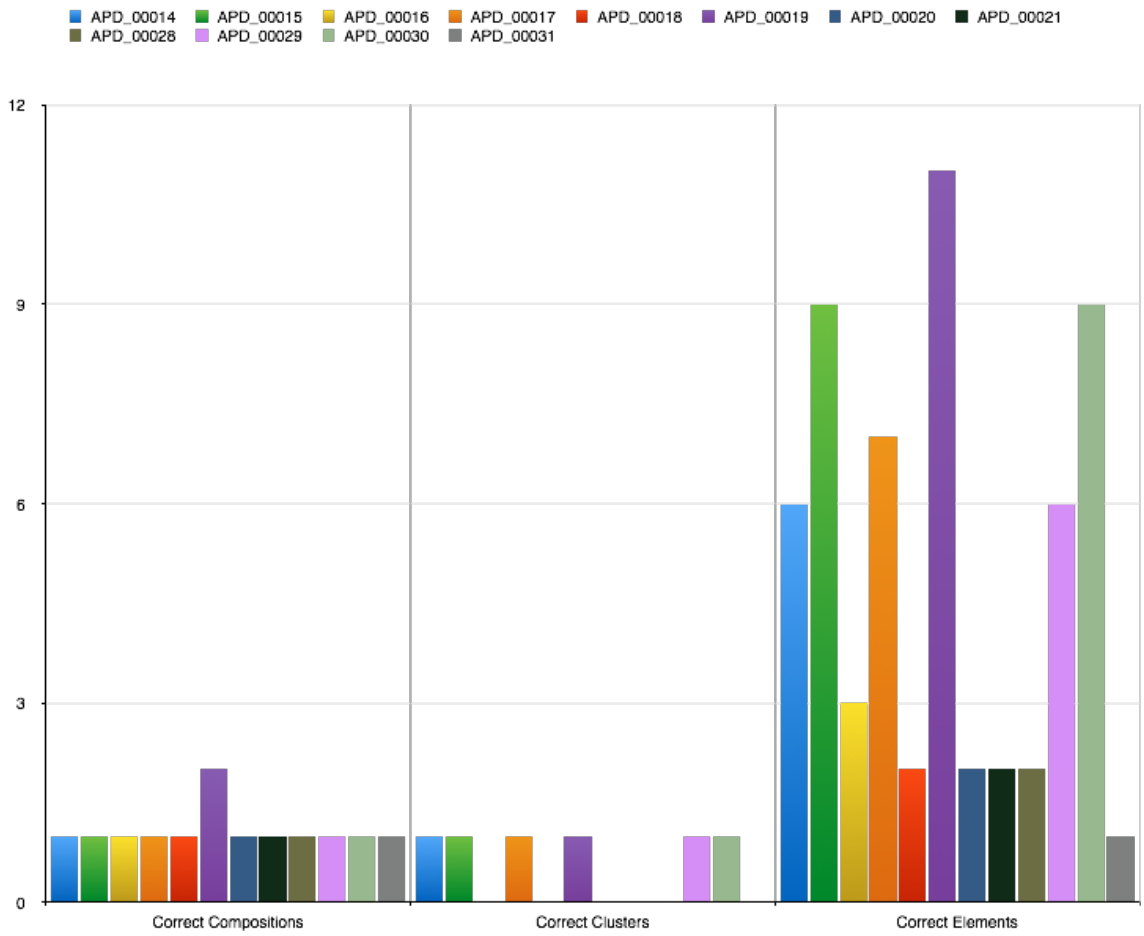
were twelve cases of understanding and one where expectations were exceeded, however there was a higher proportion of misunderstanding due to participant error, in this case fifteen examples, and one where the participant misunderstood what was needed by misunderstanding *keibi*. In this case, the question was harder, with a larger number of expected 13606 Compositions and Elements than the previous two questions, as well as the use of Safeguards, that include Clusters of Elements specifying a Control.

Chart 6: Measure of understanding for each participant

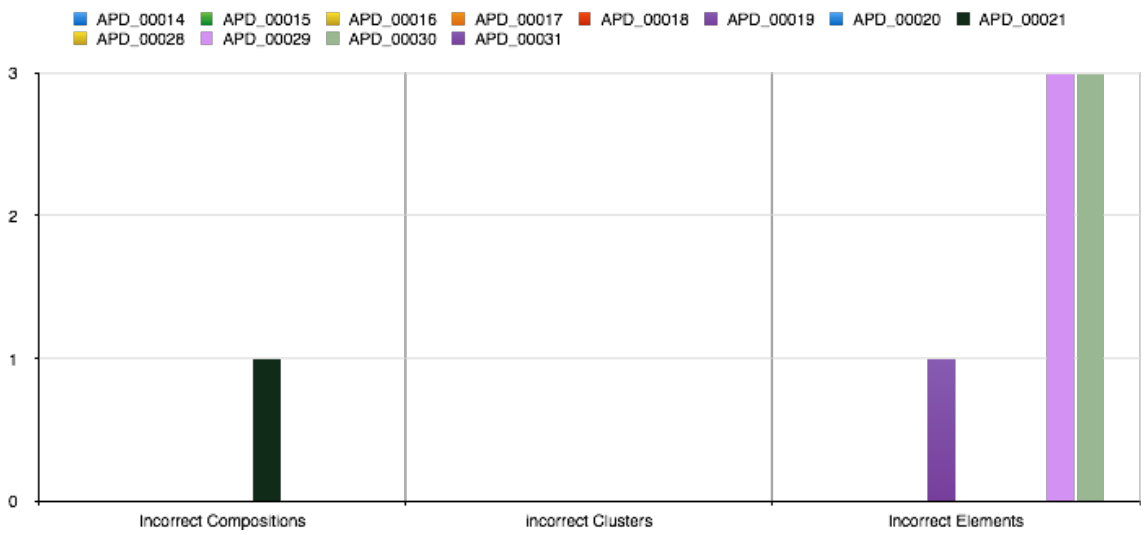
Graph 32 shows that six of the participants added between eight and fourteen Classes each, whilst the other six between two and four only. Graph 33 breaks down the number of correct Classes with mostly one Composition and in one case two Compositions, where APD_00019 was the only case that correctly added an Activity in addition to the Safeguard. Six participants added Control Clusters and a correspondingly higher number of Element items, between six and eleven in each case. The other six participants failed to add the Control Cluster and accordingly had Element numbers between one and three. Graph 34 shows that there were comparatively few incorrect elements added by three participants and numbering between one and three and one case of an incorrect Composition due to a misunderstanding of how to use *keibi*. Graph 35 shows a higher number of omissions, where each participant failed to add at least one Composition, usually the Activity of Research Data Sharing that was expected.



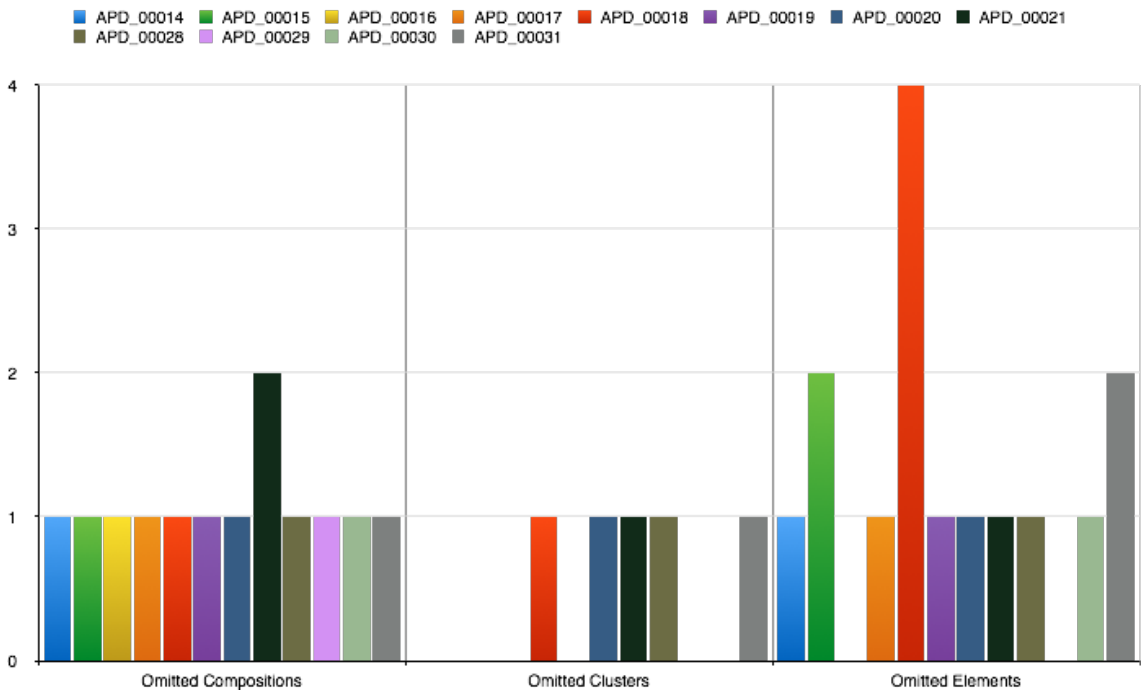
Graph 32: Number of Details added for Question 3



Graph 33: Number of correct 13606 classes added in question 3

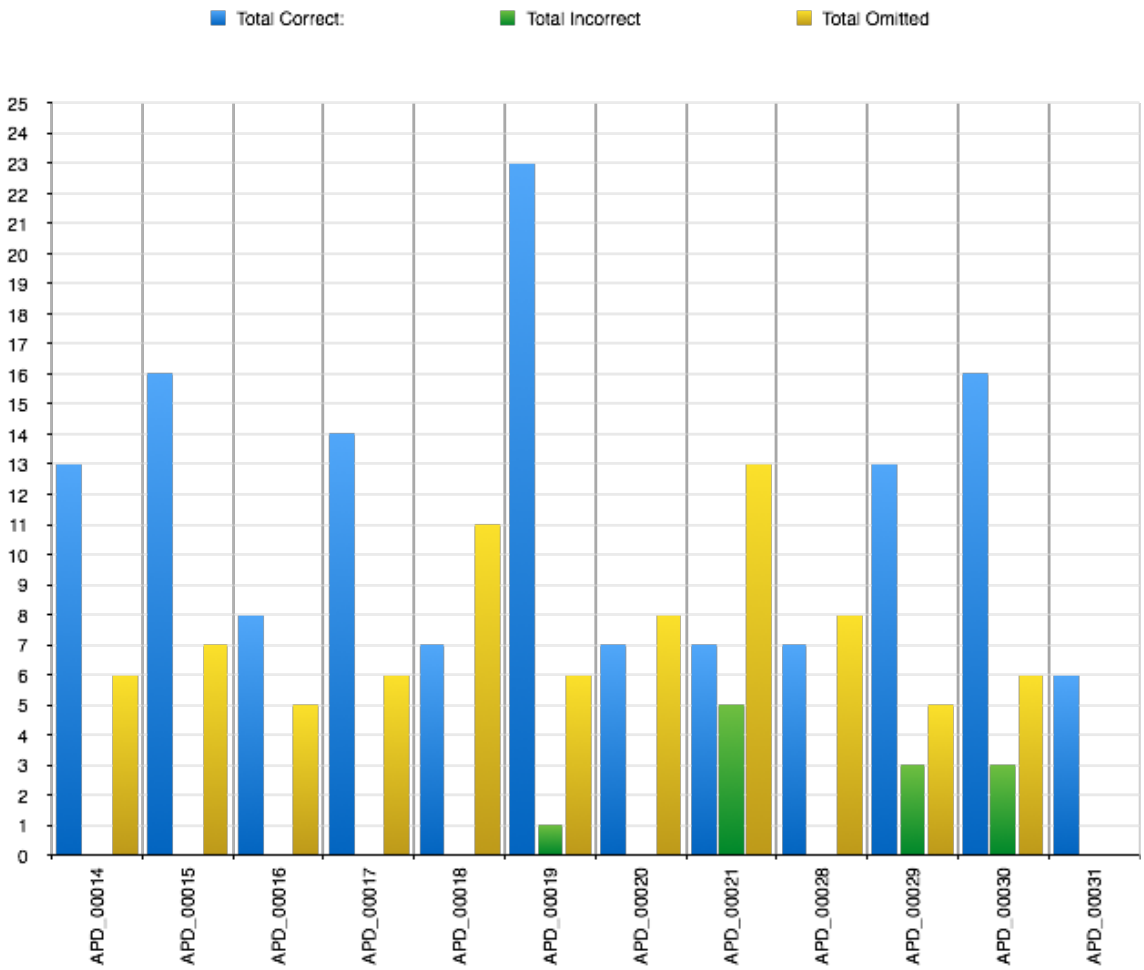


Graph 34: Number of incorrect 13606 classes added in question 3



Graph 35: Number of omitted 13606 classes

Graph 36 provides the total scores of correct, incorrect and omitted Classes, which reflect the higher proportion of omitted classes as a result of the higher levels of misunderstanding due to participant error than the previous questions. In this question, the results indicated that there were higher levels of omission than incorrect Classes, in the case of participant APD_00021 as a result of misunderstanding *keibi* itself. The higher complexity and expectation of having an Activity and Safeguard Composition added was not clear to the participants, indicating that they were still getting used to the tool and its use. There was also a high difference between the total correct scores, where the lowest was six and highest was twenty-three, directly consistent with the numbers added



Graph 36: Total Scores for Classes added in Question 3

Question 4 - Computable policy removal of identifiers:

This question was designed to provide computable policy items, where identifiable attributes were supposed to be removed before data was shared. Chart 7 shows that there were eleven cases of misunderstanding due to participant error against fourteen cases of understanding. Graph 37 shows the numbers of details that were added by the participants, showing a large variation of between three and sixteen details added. Graph 38 shows the breakdown of correct Classes added between participants, where only four added the Safeguard Cluster and only two added more than one Composition. There is a sizeable variation in the number of Elements correctly added, between two and nine in total. Graph 39 shows there were again comparatively few errors, though Graph 40 shows many more

omissions – nearly all participants omitted at least one Composition, Cluster or Element.

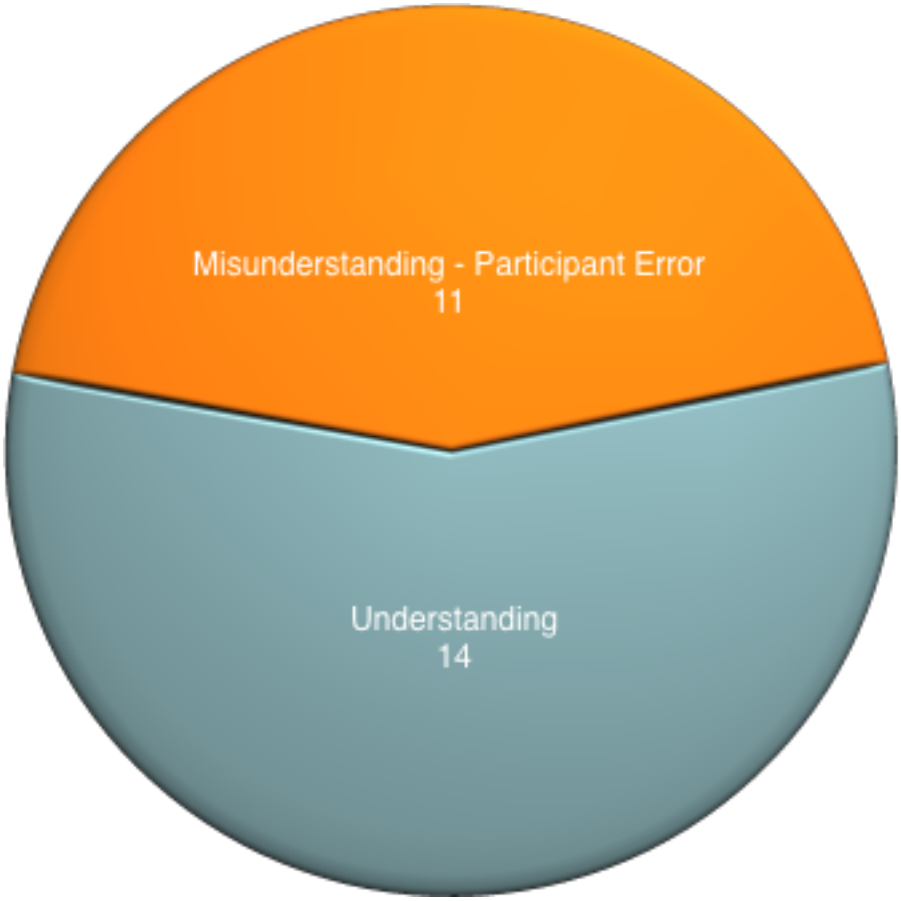
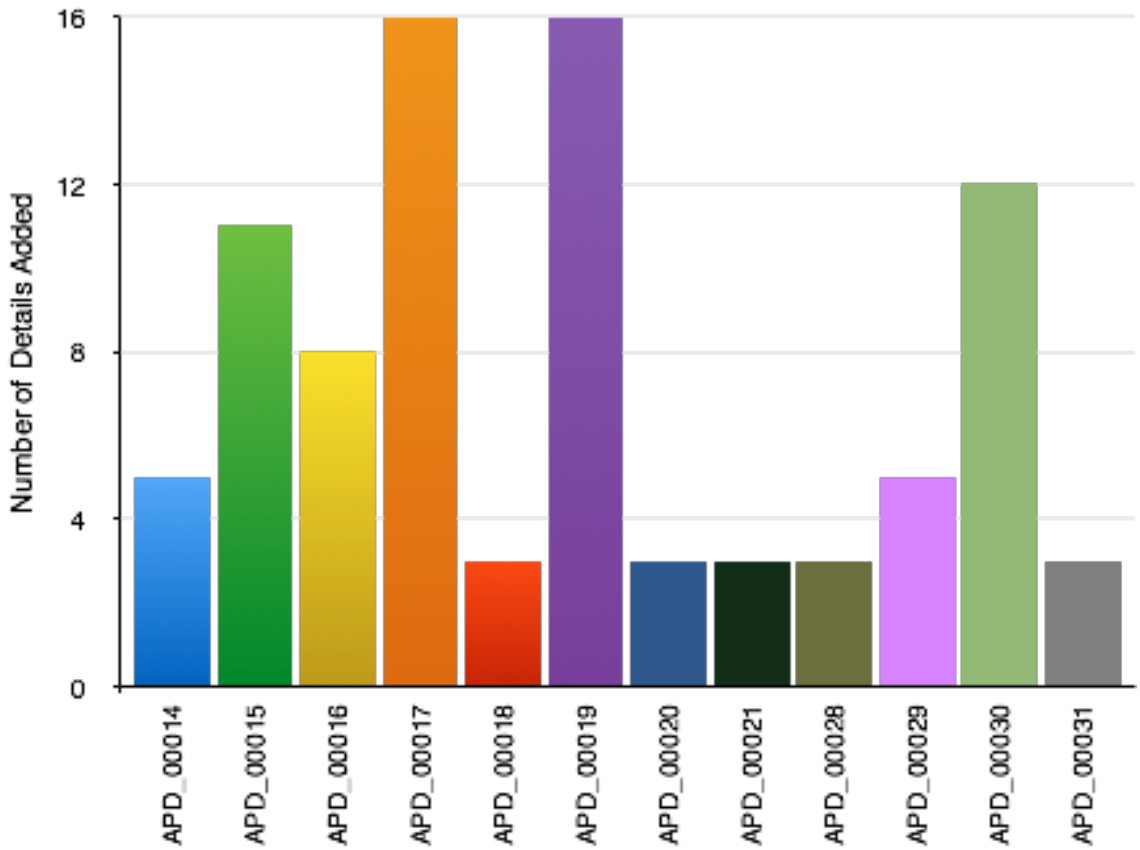
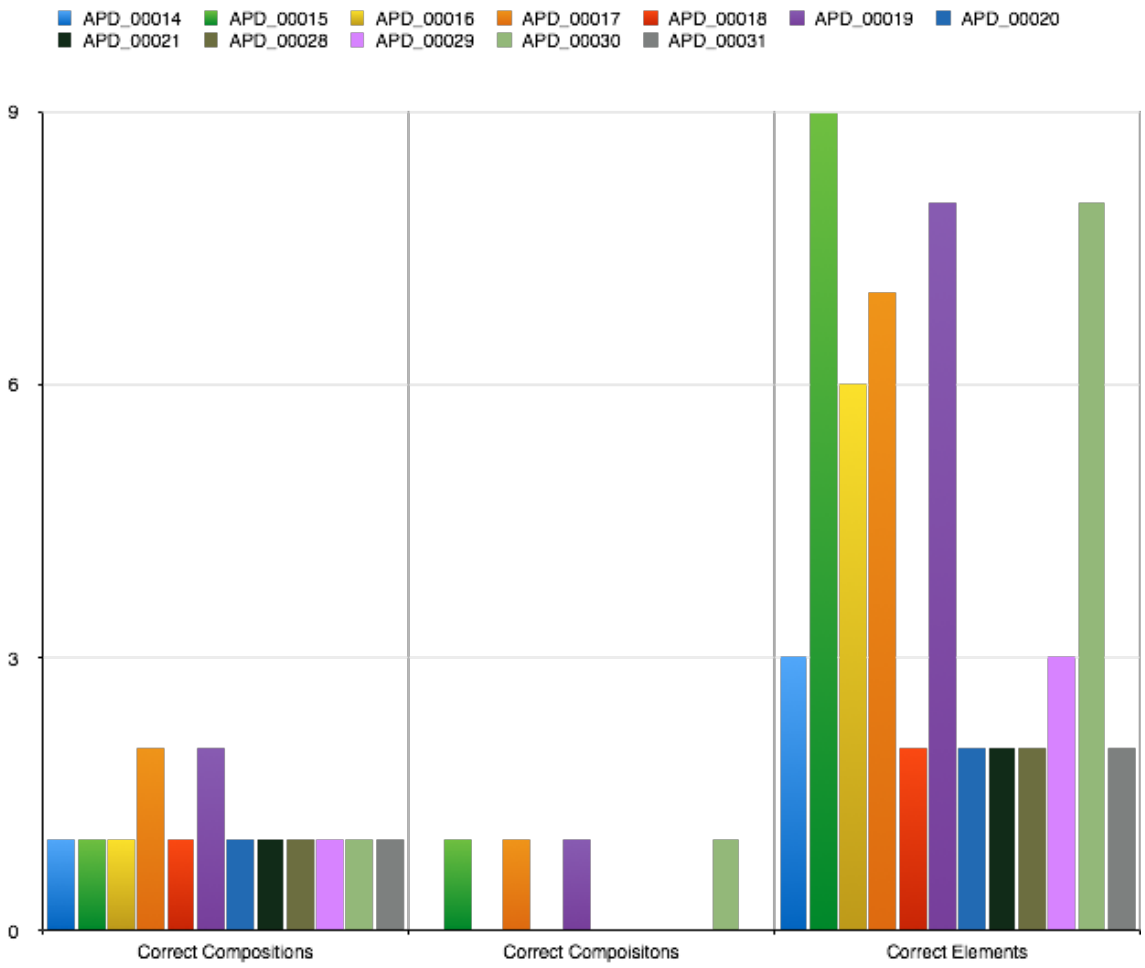


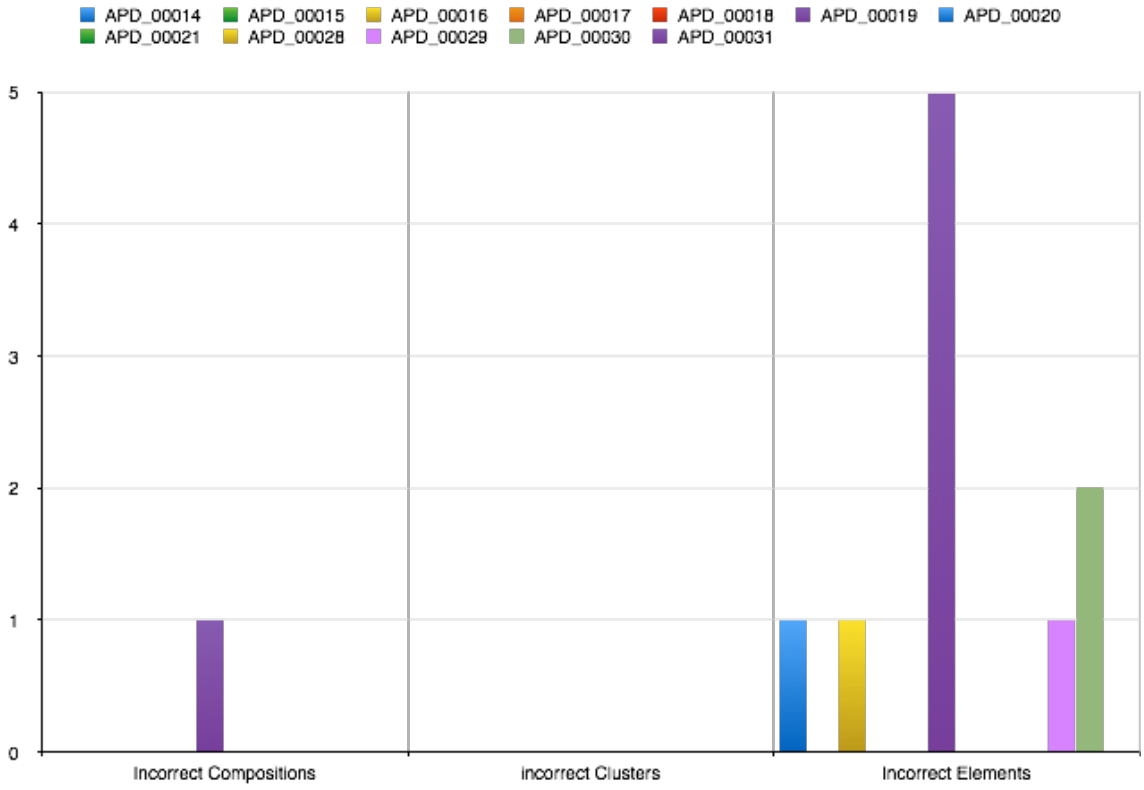
Chart 7: Proportion of understanding for each participant in question 4



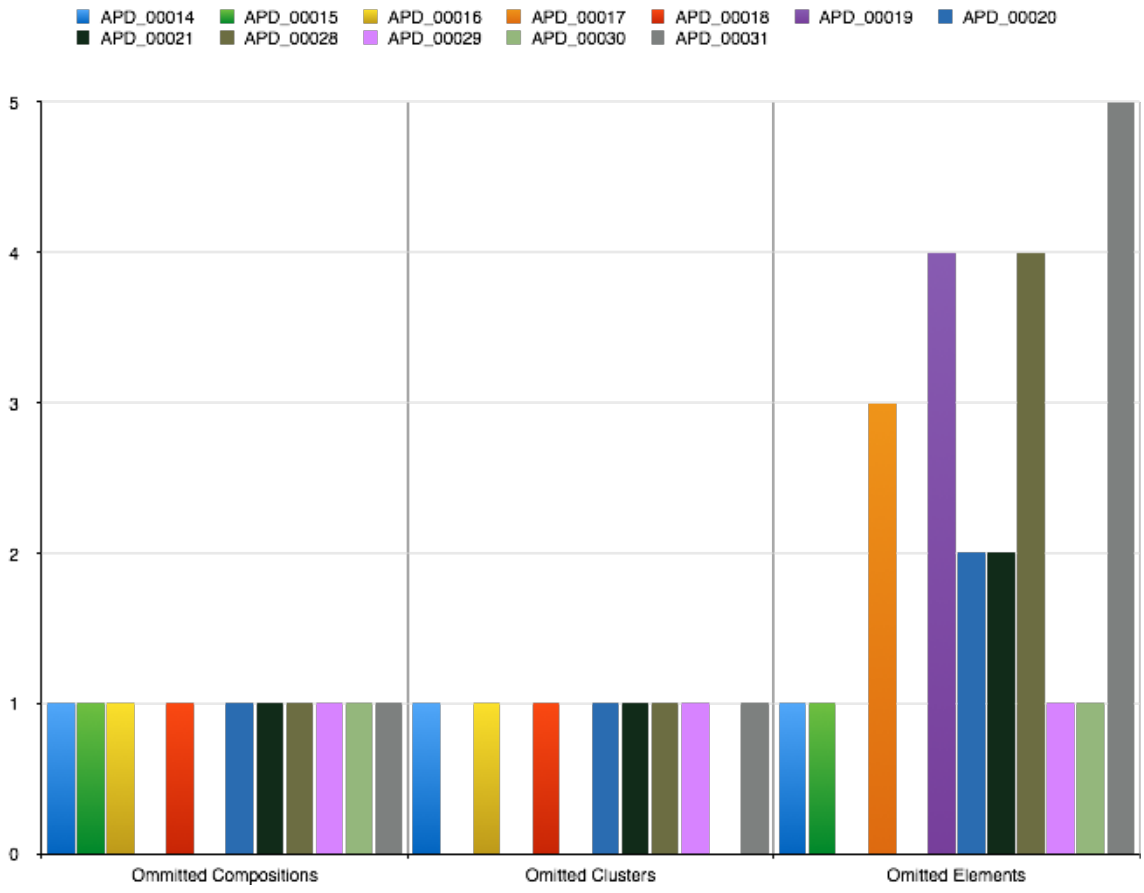
Graph 37: Number of Details added for Question 4



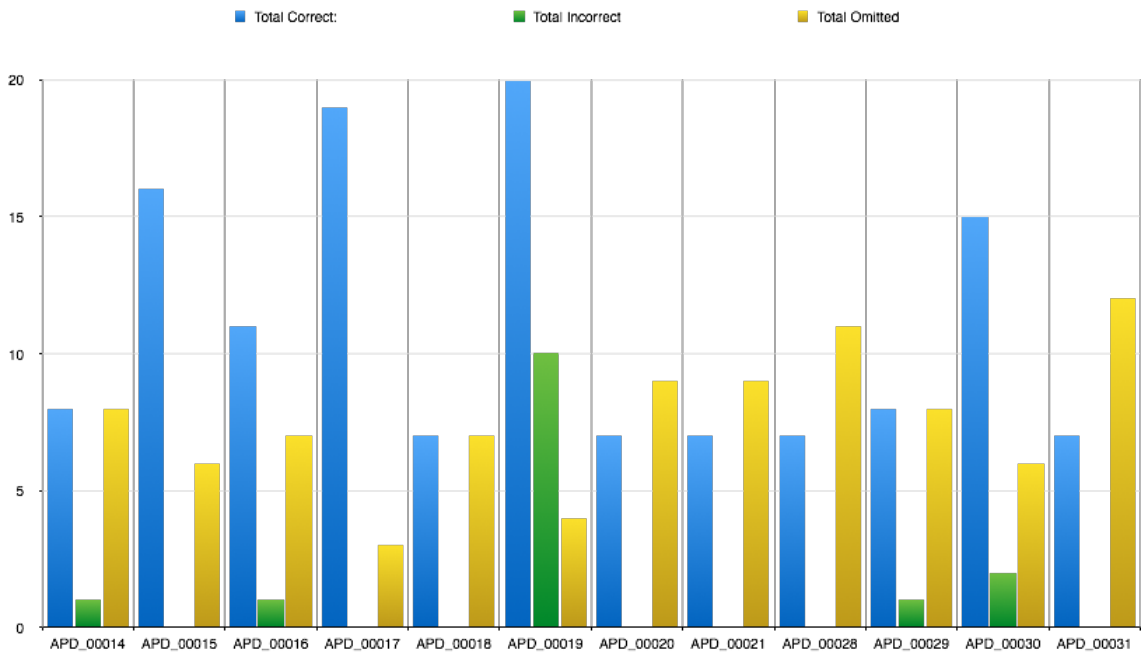
Graph 38: Number of correct 13606 classes added in Question 4



Graph 39: Number of incorrect 13606 classes added in question 4



Graph 40: Number of omitted 13606 classes in question 4



Graph 41: Total Scores for Classes added in Question 4

Graph 41 shows the total scores across participants for correct, incorrect and omitted Classes. The results seem consistent with those for question 3: the proportion of omitted to correct classes appears to be higher than incorrect classes; there is also a variation between the total correct scores, the lowest being seven and highest twenty. The profile of total scores suggests that participants were still getting used to the tool and growing their confidence with it.

Question 5 Sheet 1: Consent Requirements for Research Data Sharing

This question again showed some participant misunderstanding, relating to errors but also omissions in specifying Safeguards, illustrated by Chart 8, but also a case of exceeding expectations. Graph 10 shows the total number of Classes added, where the maximum was thirty-five and minimum ten.

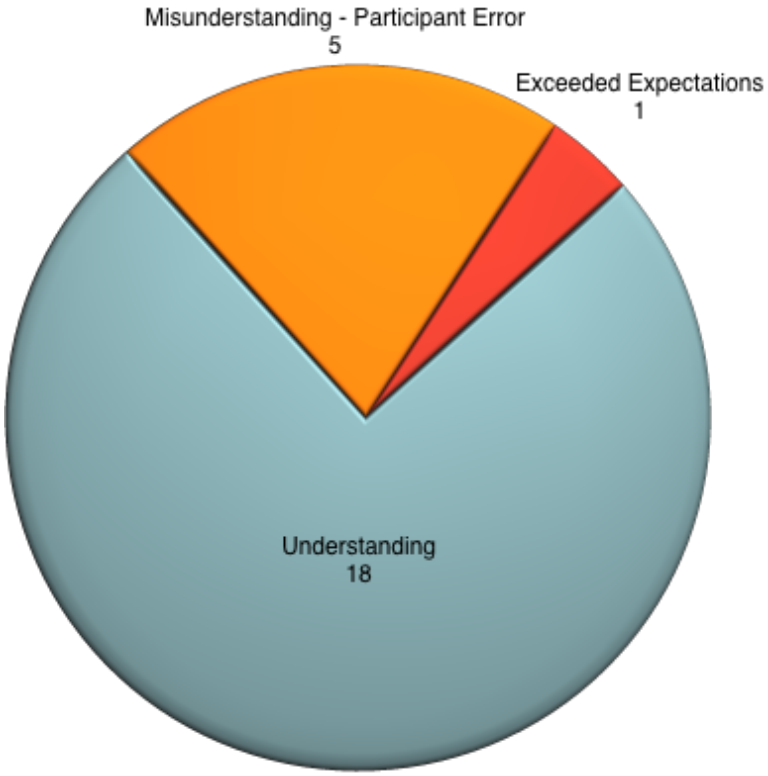
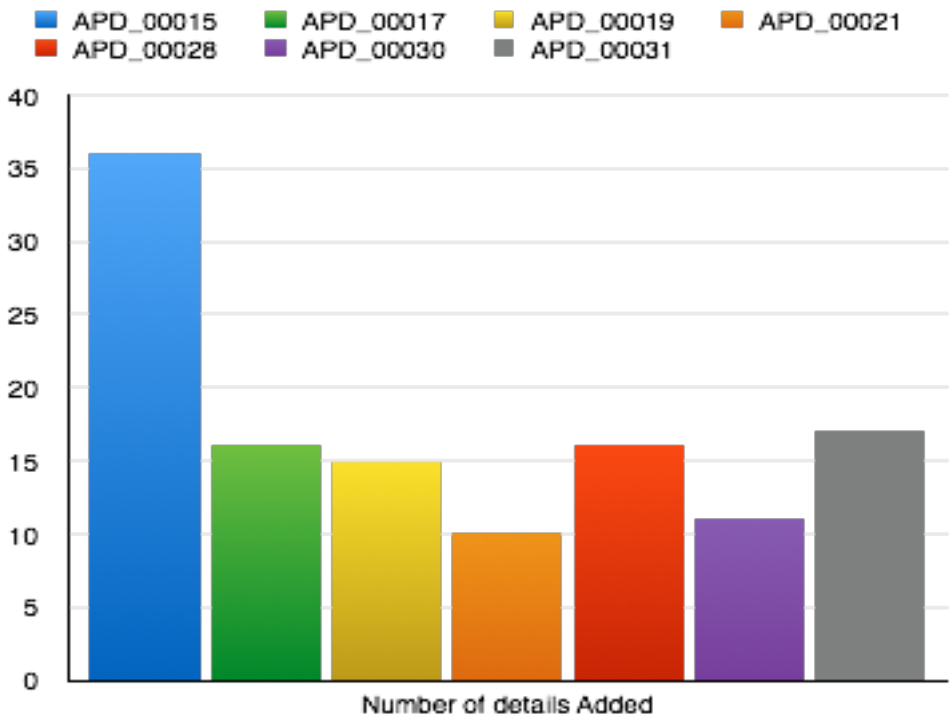


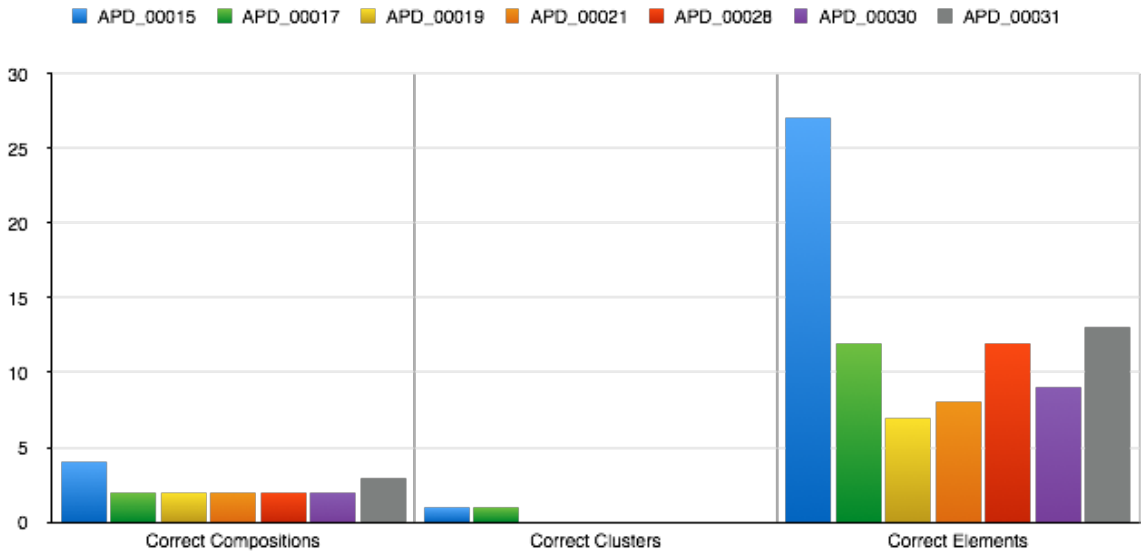
Chart 8: Measure of understanding for each participant



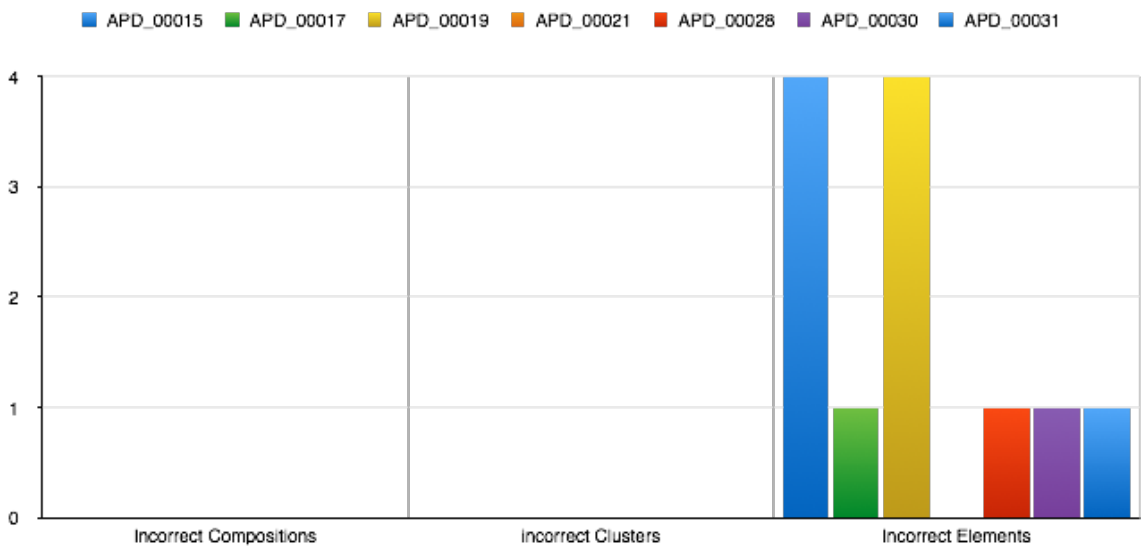
Graph 42: Number of Details added for Question 5 Sheet 1

Graph 43 shows the number of correct Classes added, where there is considerable variation between the use of clusters (where two participants added one and the

rest none. There is also a variation of number of correct elements, where there were a maximum of twenty-seven and minimum of seven across all the participants. Particioant APD_00015 had a significantly higher number of Elements than the other due to adding two Safeguard Compositions, and one Control Cluster.

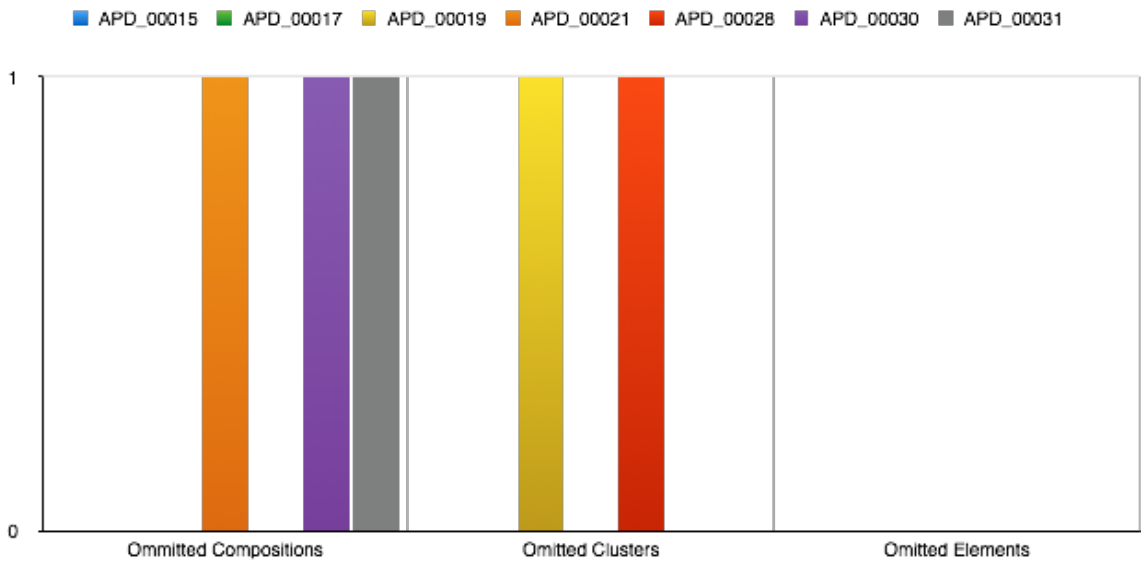


Graph 43: Number of correct 13606 classes added



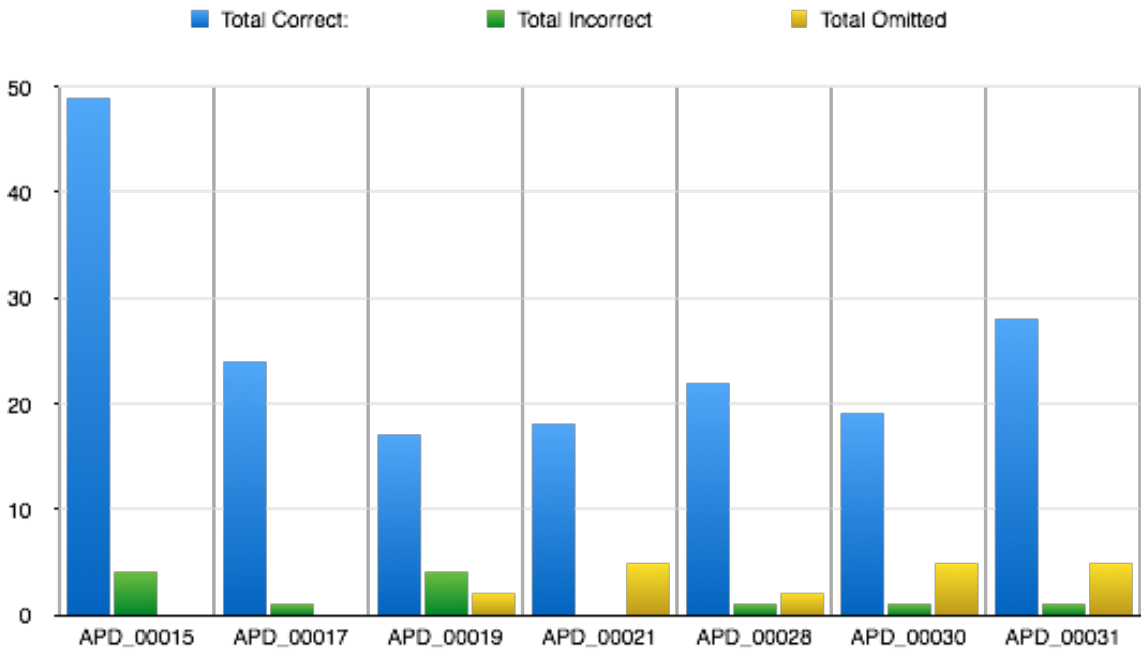
Graph 44: Number of incorrect 13606 classes added

Graph 44 shows the number of incorrect classes that were added, in this case, only Elements. There was some evidence here of participants embellishing the details that they added, sometimes incorrectly, but this does show that they were becoming bolder with answering the questions and attempting to understand the wider context within which the Safeguards were intended to operate.



Graph 45: Number of omitted 13606 classes

Graph 45 shows the omission of certain Activities from previous questions and some omission of Controls again. Omissions were also more common than error in this case. Graph 46 illustrates this, showing the total correct, incorrect and omission scores.



Graph 46: Total Scores for Classes added in Question 5 Sheet 1

Question 6 Sheet 1 - Mac Mini Networking

The results from this question show a higher proportion of misunderstanding due to participant error than in previous questions and two cases of exceeding expectations, as shown in Chart 9.

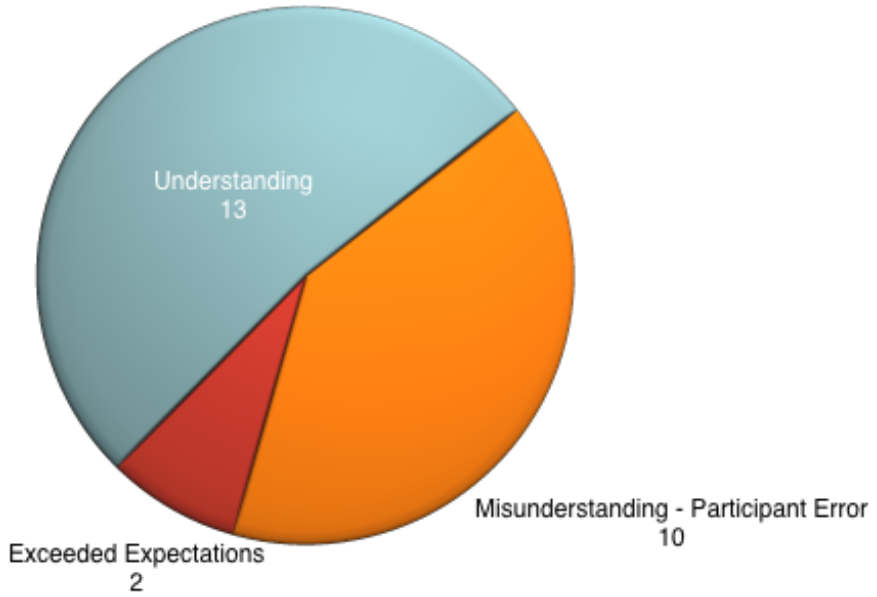
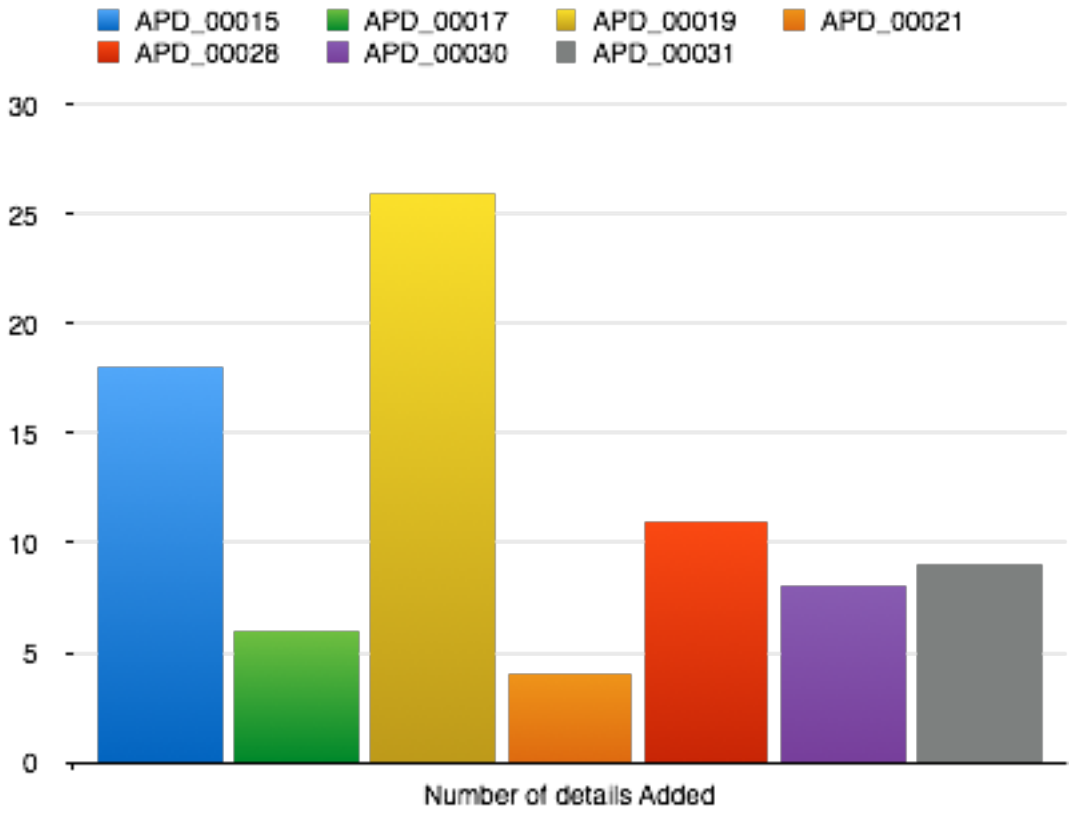
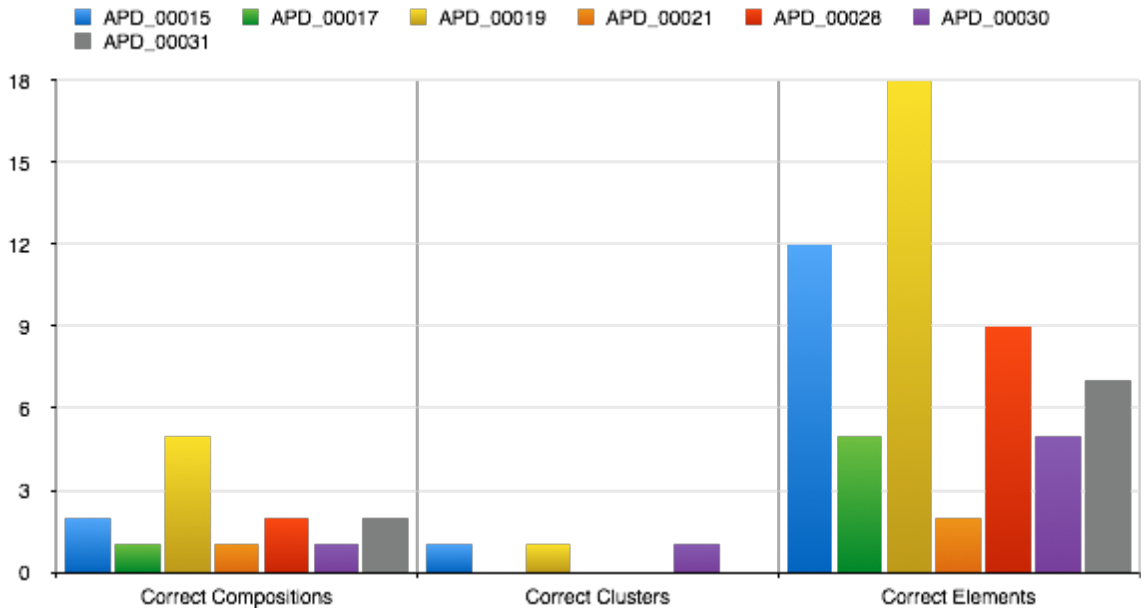


Chart 9: Measure of understanding for each participant

Graph 47 shows the total number of Classes added, where the highest number was twenty-six and the lowest four. This variation is reflected in the number of correct elements that were added, the highest being eighteen and the lowest two. This variation can be explained by the number of omissions that occurred



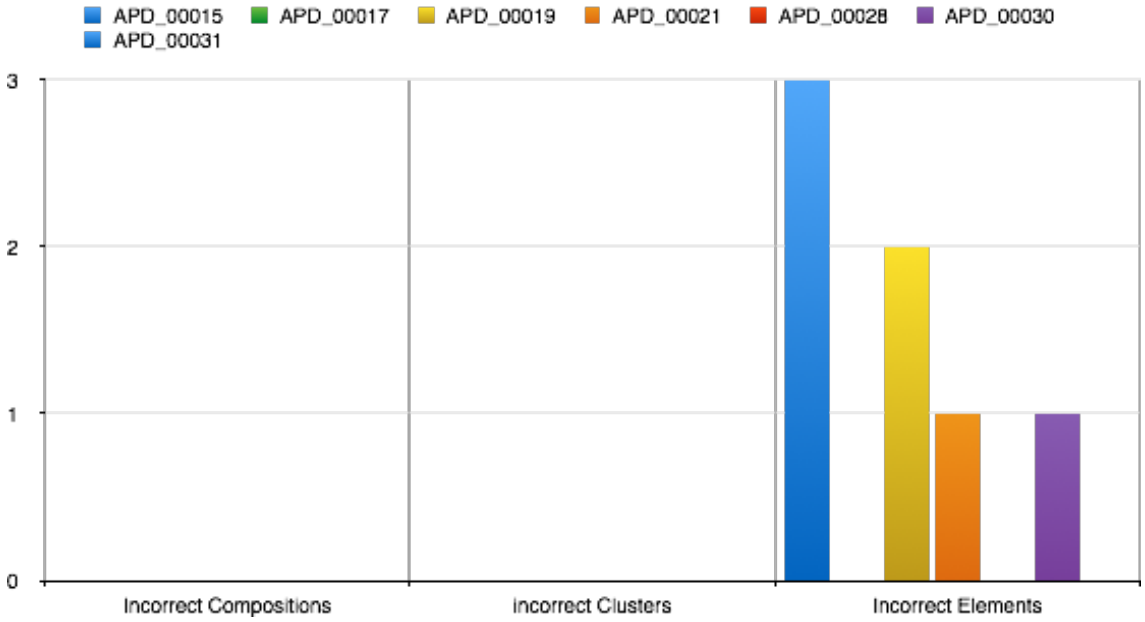
Graph 47: Number of Details added for Question 6 Sheet 1



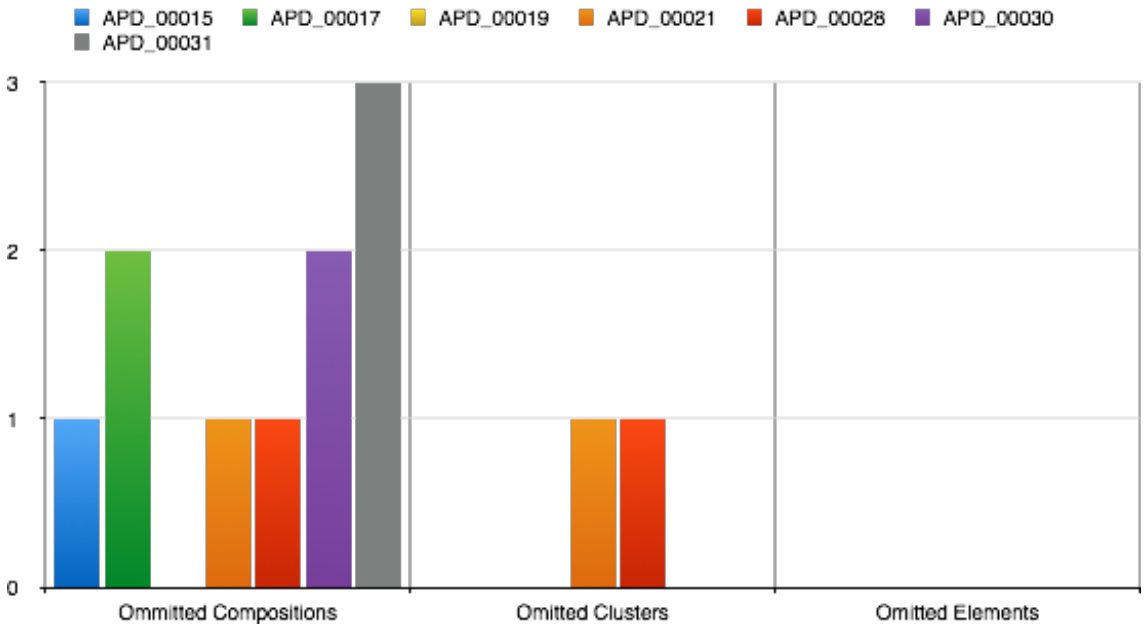
Graph 48: Number of correct 13606 classes added

Graph 48 shows that there few errors, mostly relating Element items that were relatively minor, though Graph 49 shows an explanation for the higher level of

participant misunderstanding and low numbers of details added, which were due to a high number of Composition and Cluster omissions as shown in Graph 50.



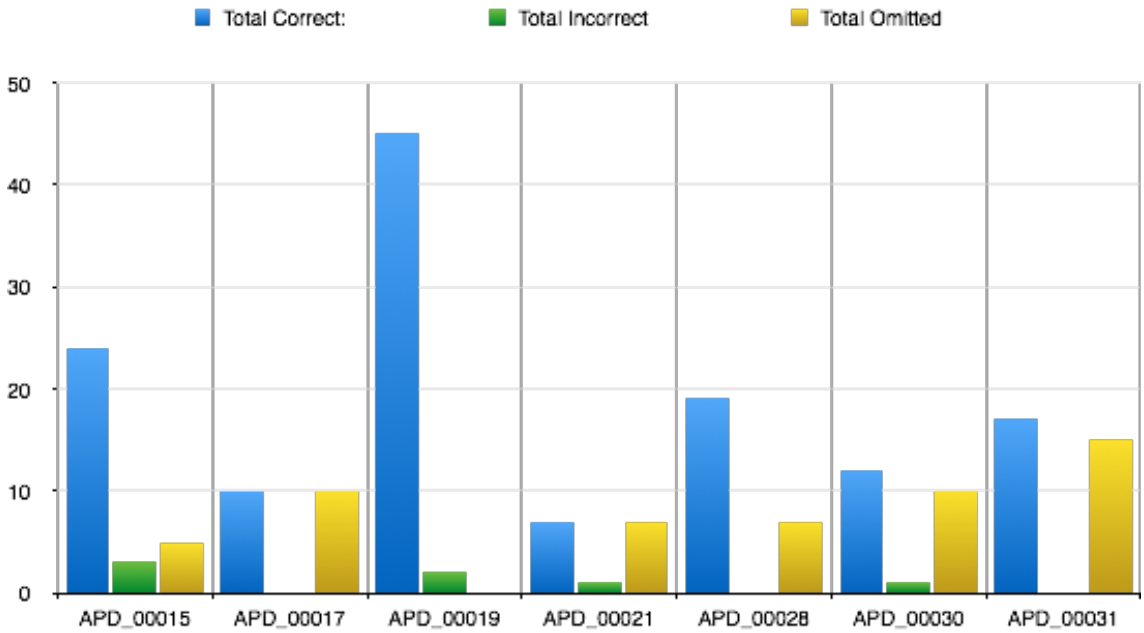
Graph 49: Number of incorrect 13606 classes added



Graph 50: Number of omitted 13606 classes

Graph 51 shows the total scores for correct, incorrect and omitted EN 13606 Classes. This shows a higher proportion of omissions, in this case for participant APD_00030 and APD_00031 in this question. The omissions related mostly to the

PDA Information Asset, where only one participant added it, though there were some Safeguard Omissions as well.

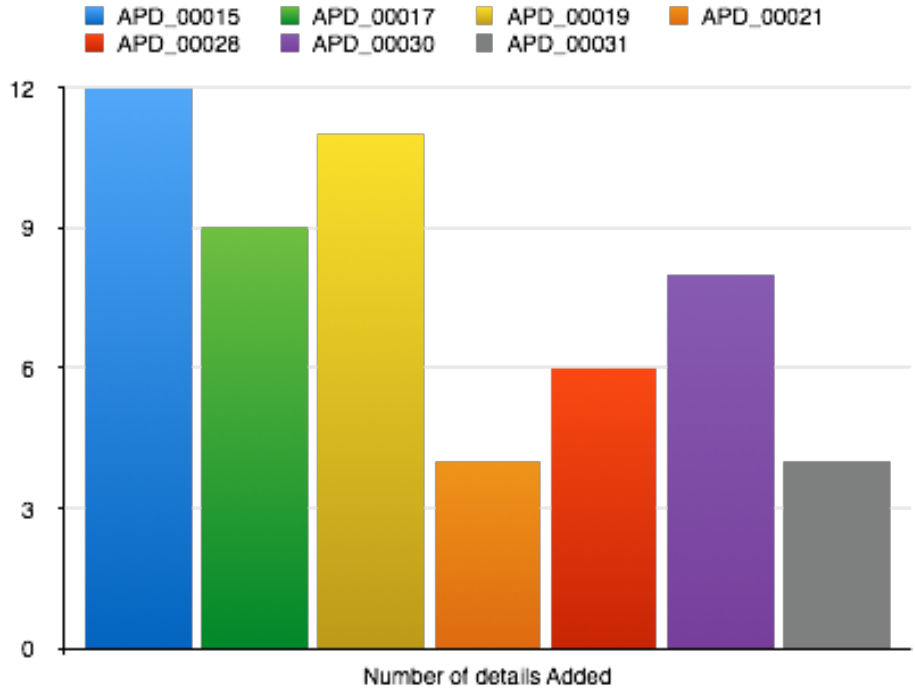


Graph 51: Total Scores for Classes added in Question 6 Sheet 1

Question 7 Exercise Sheet 1 - Sharing data from the Mac Mini

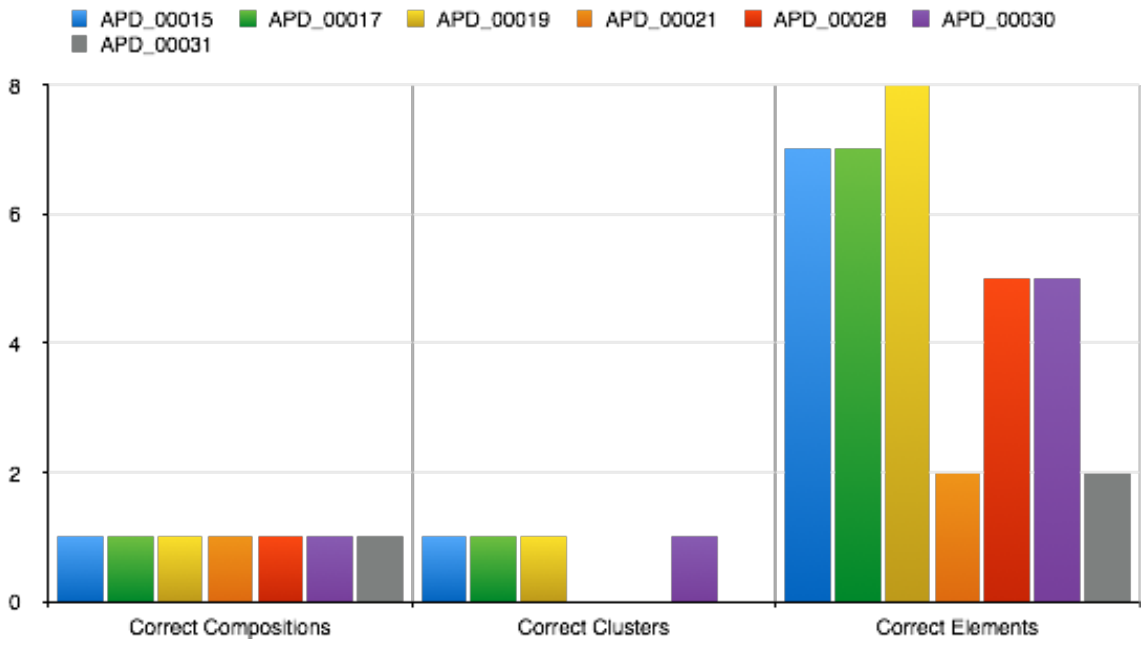
The responses to this question showed a seventy per cent understanding score, as demonstrated in Chart 10. Graph 52 shows that there was still some variation in the number of classes between the maximum and minimum, four being the minimum and twelve being the maximum.

Chart 10: Measure of understanding for each participant

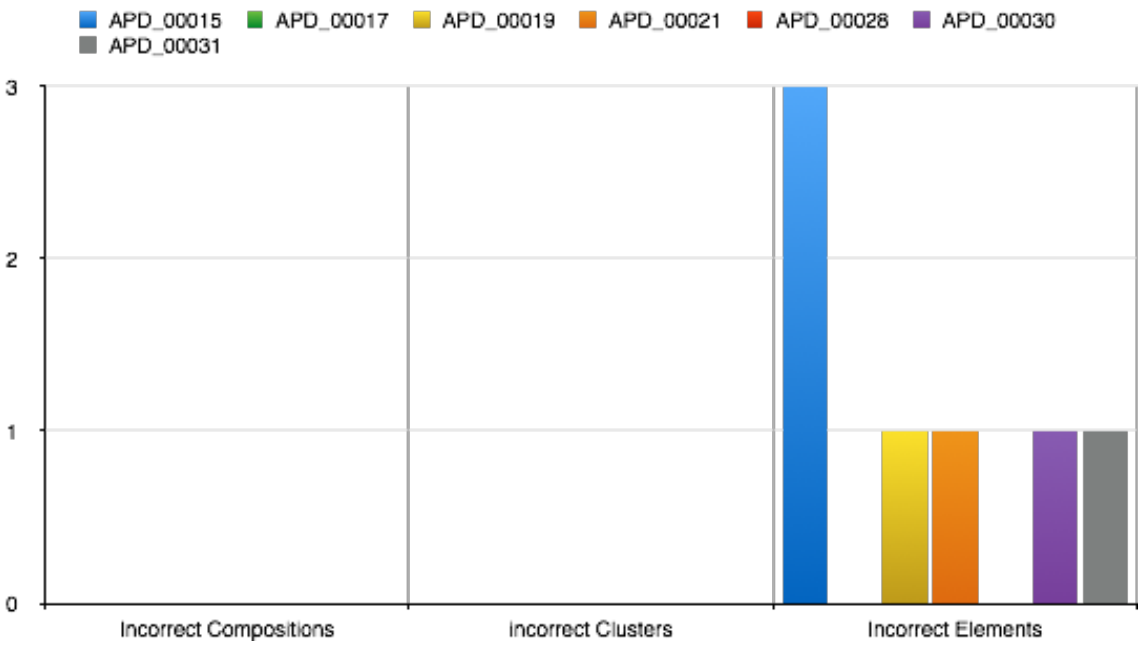


Graph 52: Number of Details added for Question 7 Sheet 1

Graph 53 shows the number of correct classes added for this question, where four out of the seven participants added a Control Cluster, which suggests that participants had reached a point in the exercise where they were more familiar with the features of the Secutype Model and *keibi*. The lowest number of elements was two and highest eight, showing some variation.

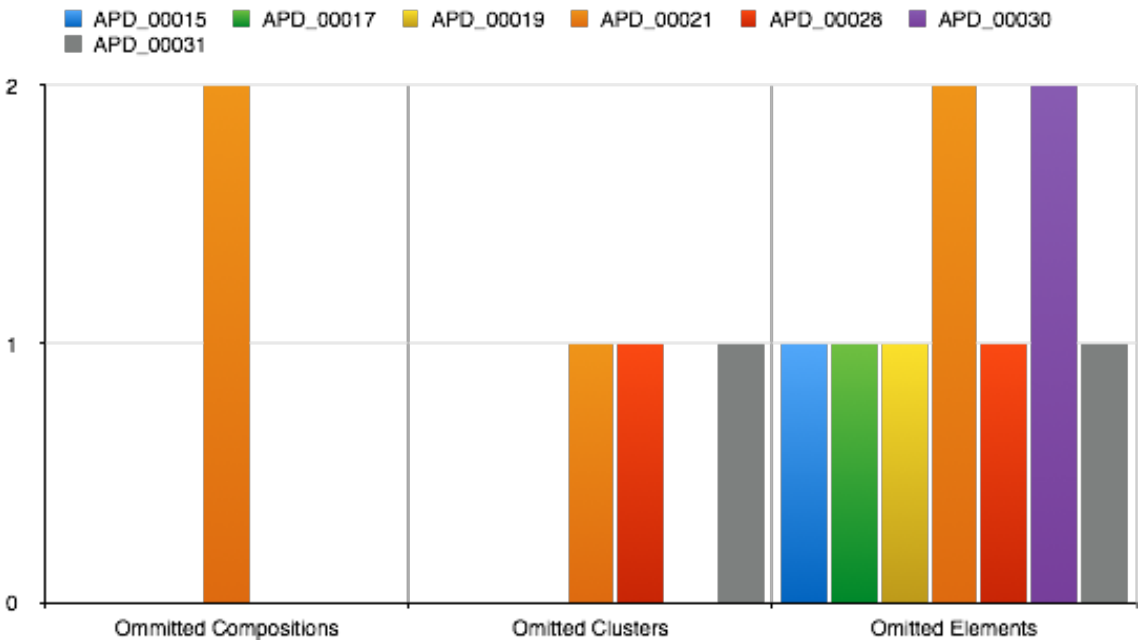


Graph 53: Number of correct 13606 classes added

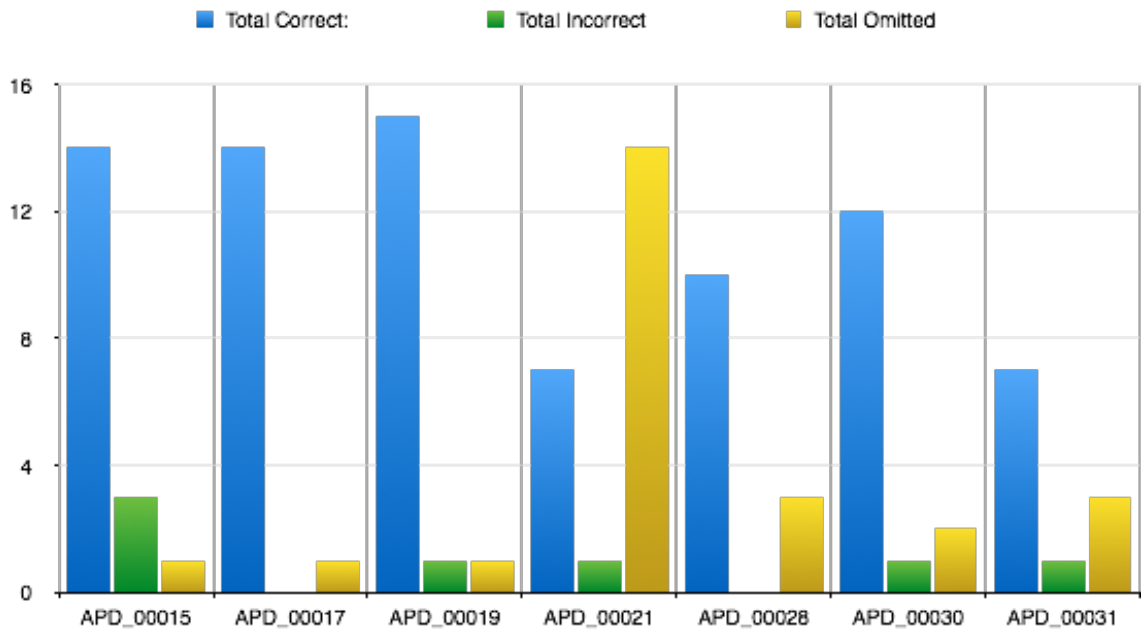


Graph 54: Number of incorrect 13606 classes added

Graph 54 shows the number of incorrect Elements, which were fewer, though there was some omission: one participant omitted a Composition and three a Cluster each. Graph 55 shows that every participant omitted at least one element, and in two cases, two each. The participants tended to omit a reference to the CD-ROM Information Asset.



Graph 55: Number of omitted 13606 classes



Graph 56: Total Scores for Classes added in Question 7 Sheet 1

Graph 56 shows the total scores across all the participants for each of the participants. Participant APD_00021 had a high number of omissions, mainly due to their being frustrated by the system use related error made in question xxx, where they had authored an entire Safeguard, but failed to save it before navigating away from the page.

Question 5 Exercise Sheet 2 - Audit

Chart 11 shows the distribution of understanding, misunderstanding and exceeding expectations. They exceeded expectations three times, where one of those cases was to add a Legal Basis. They misunderstood as a result of their own error three times and showed twelve cases of understanding how to use *keibi* to author policies correctly. Graph 57 shows the number of details added for each participant, where the lowest number was eight and the highest was thirty-two.

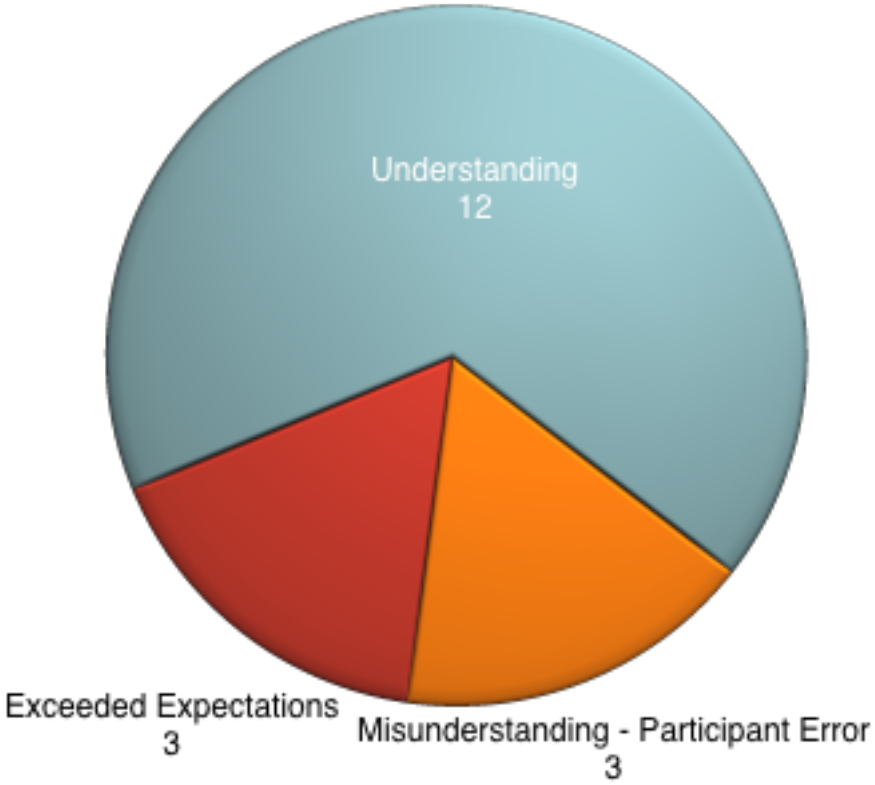
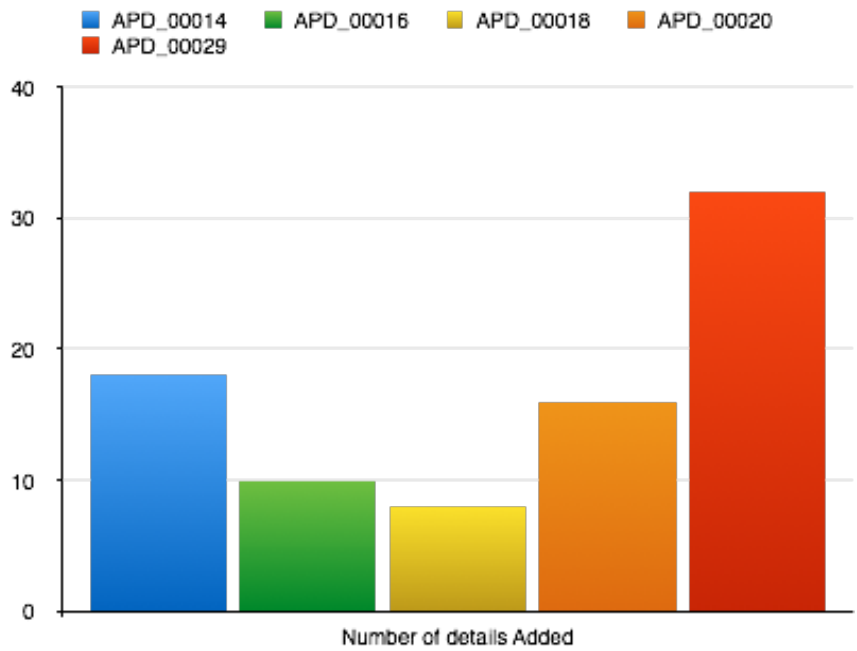


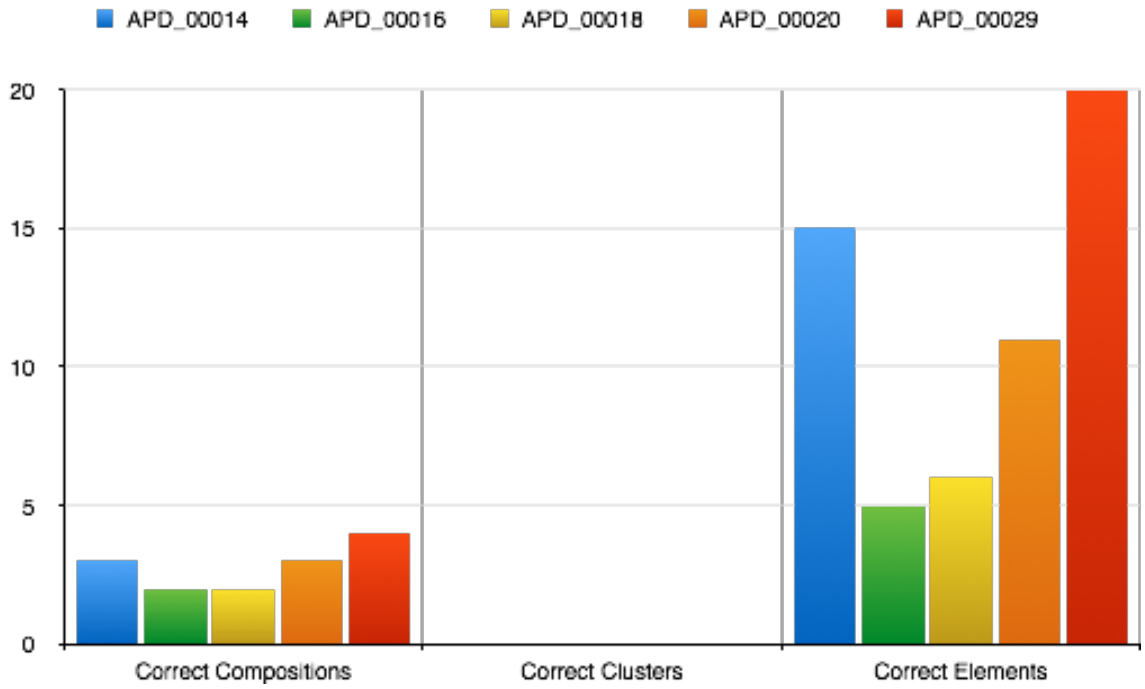
Chart 11: Measure of understanding for each participant



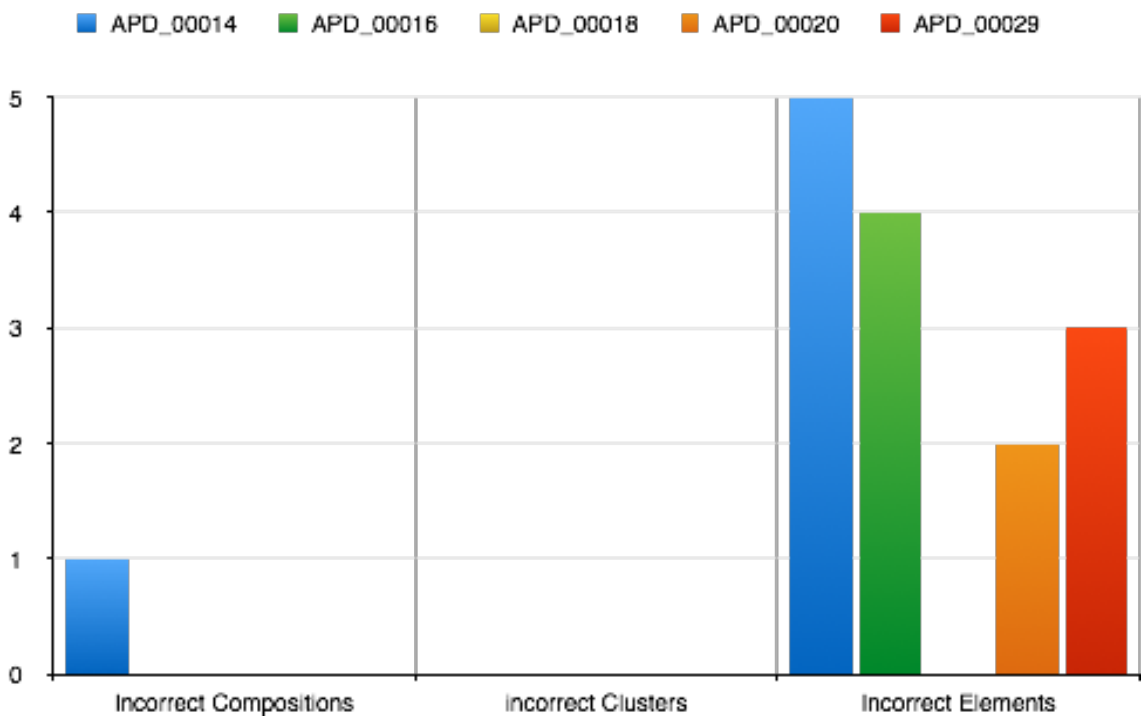
Graph 57: Number of Details added for Question 5 Sheet 2

Graph 58 provides the total numbers of correct Classes added. None of the participants added any Control Clusters, and there was a variation in numbers of Elements added (the fewest number of five and greatest twenty added). Both

participants APD_00014 and APD_00016 added an Activity of Audit instead of a Safeguard. Whilst both are not necessarily incorrect, these have been recorded as omissions.

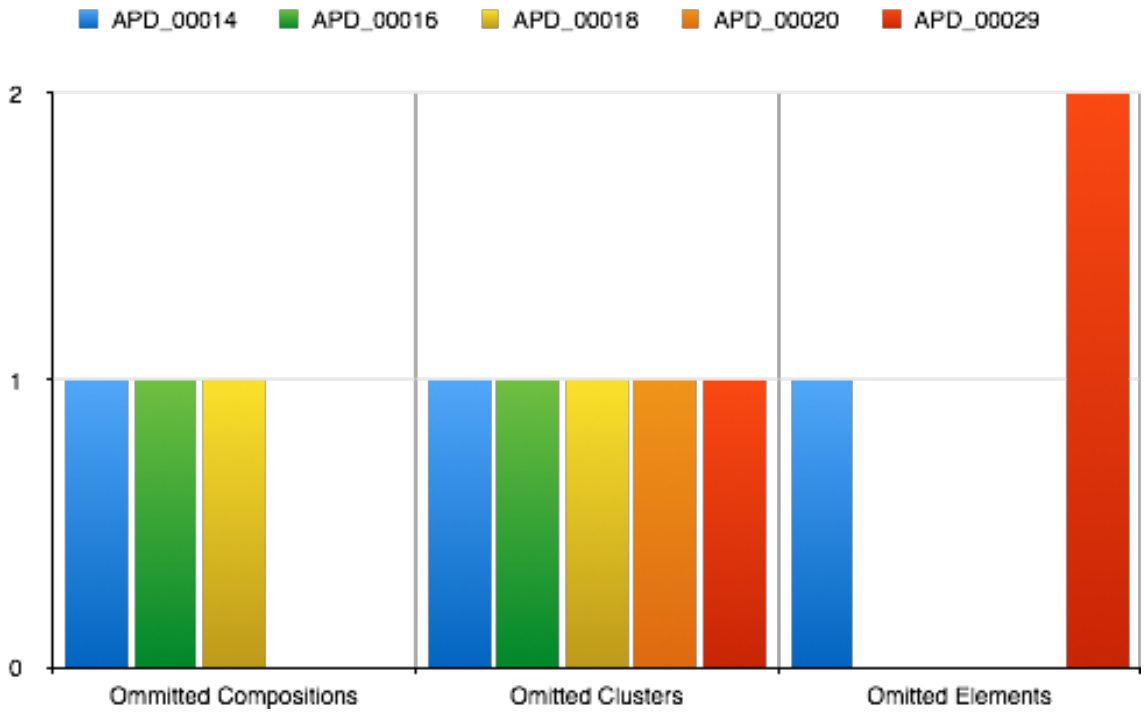


Graph 58: Number of correct 13606 classes added



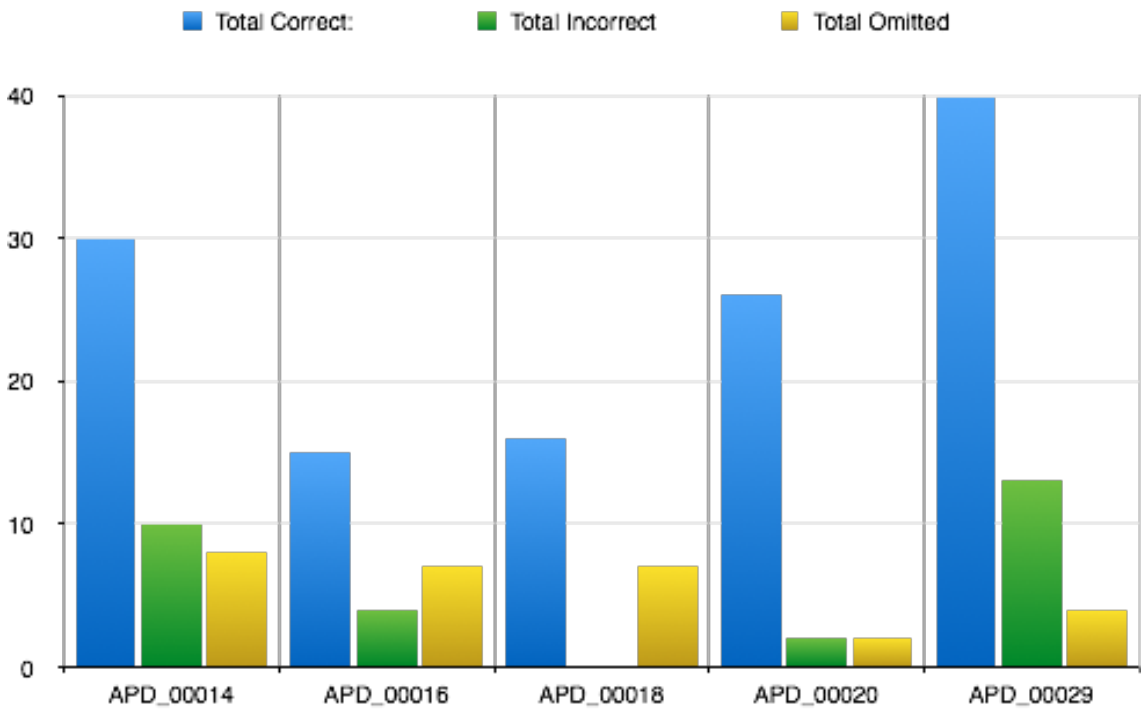
Graph 59: Number of incorrect 13606 classes added

Graph 59 shows the number of incorrect of Classes that were added across all participants. In this case, only one participant entered and incorrect Composition, with five incorrect Elements. All but one of the other participants added an incorrect Element, where errors were minor.



Graph 60: Number of omitted 13606 classes

Graph 60 shows the number of omissions across the participants. Three omitted a Safeguard and all of them omitted the Control Clusters. The responses to this question showed a higher proportion of omitted Classes to errors.



Graph 61: Total Scores for Classes added in Question 5 Sheet 2

Graph 61 shows the total scores across correct, incorrect and omitted clusters. This showed a general trend for omissions to be more prevalent than errors, where Safeguard Compositions and Control Clusters tended to be responsible for the high proportion of omissions.

Question 6 Exercise Sheet 2 - Use of Known Telephone Numbers

Chart 12 provides the proportions of understanding, misunderstanding and the exceeding of expectations across participants. Almost a third of the responses appeared to misunderstanding in this question.

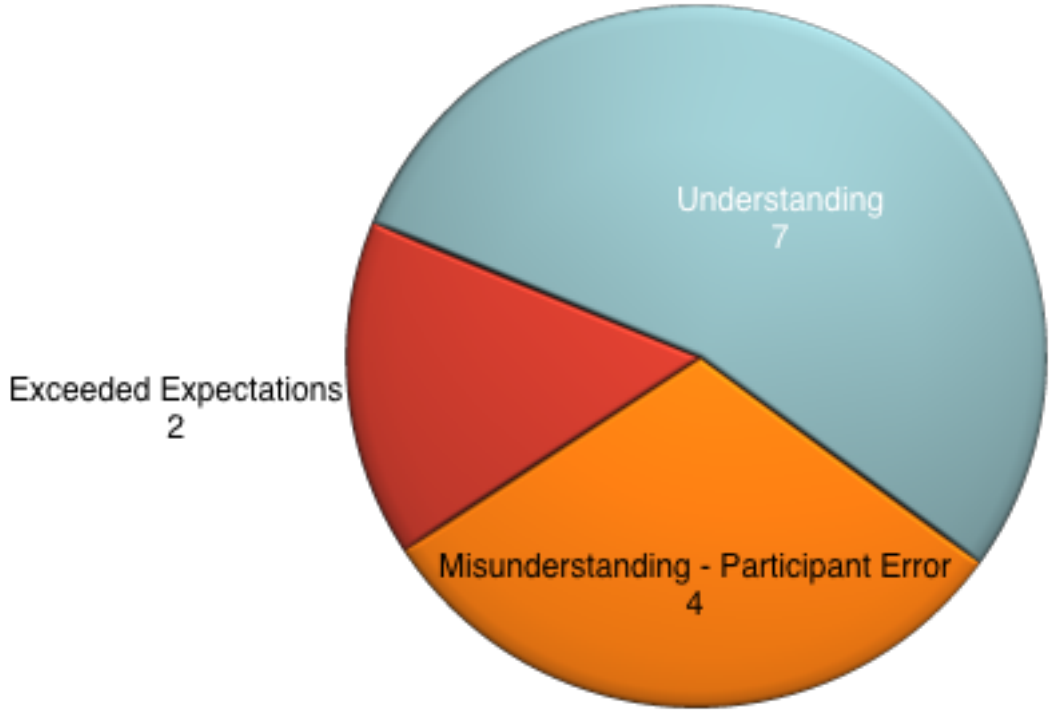
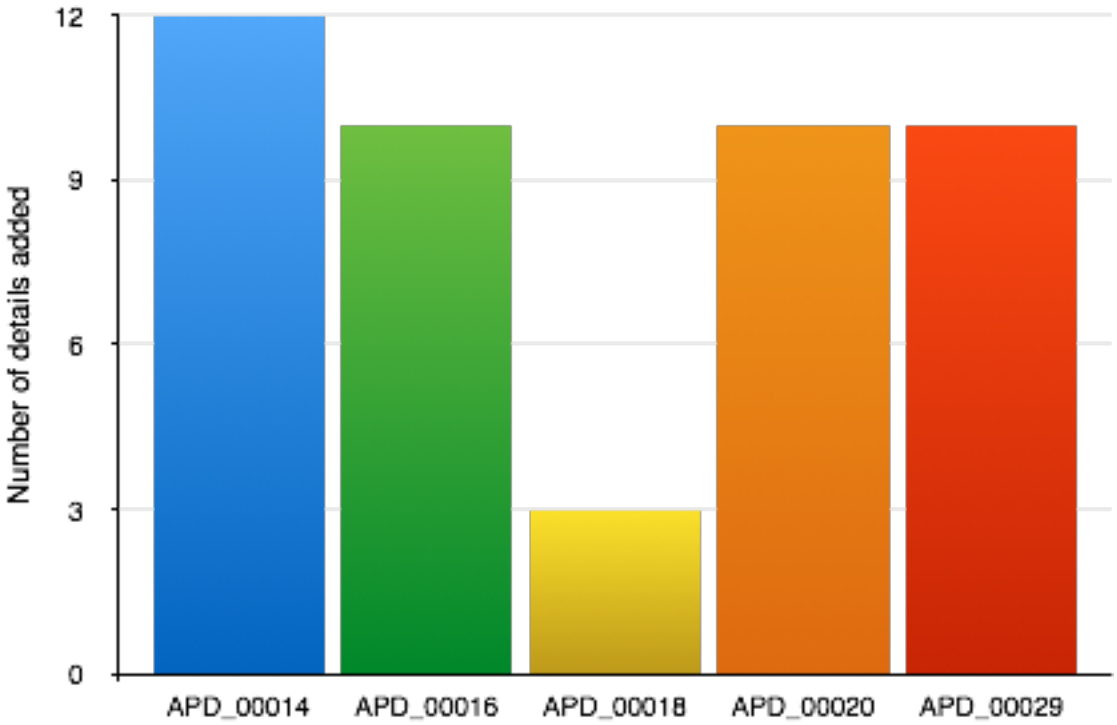


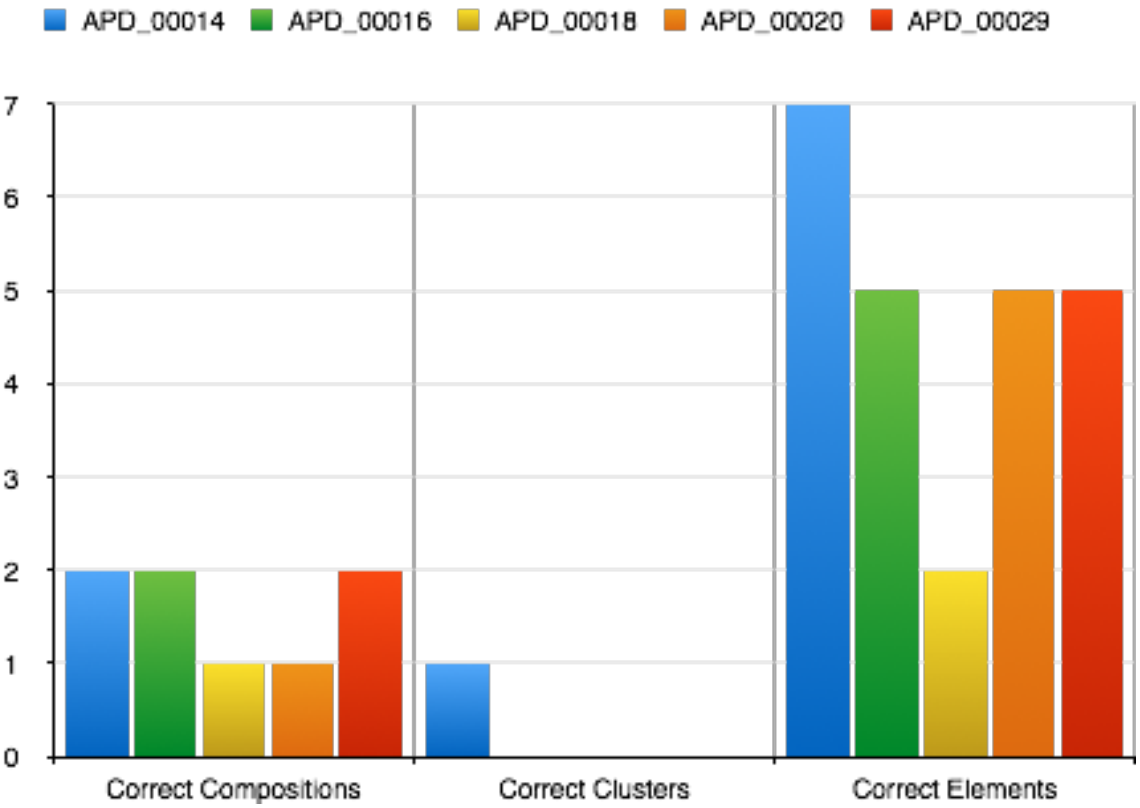
Chart 12: Measure of understanding for each participant

Graph 62 shows a degree of variation in the number of Classes added across the participants, with three being the lowest number and twelve the highest, where participant APD_00018 seemed to provide the lowest number in this question, as wit hthe previous questions. Otherwise three of the participants provided ten Classes each.

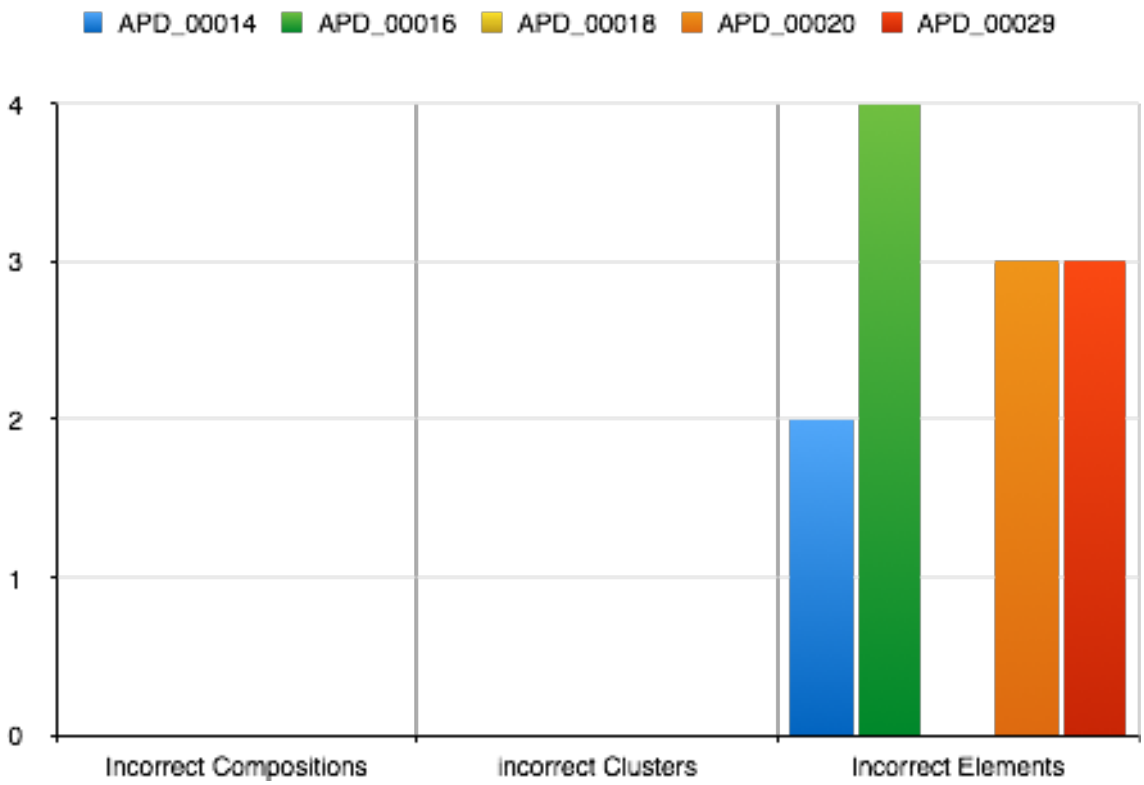
Graph 63 shows the number of correct Classes added, where only one participant added a Cluster, whilst the rest omitted them as shown by Graph 65. All participants except APD_00016 omitted an Information Asset Composition, and all except for APD_00018 made small errors in their Elements (see Graph 65). Graph 64 shows there were cases of incorrect elements across all participants except one, numbering between two and four each.



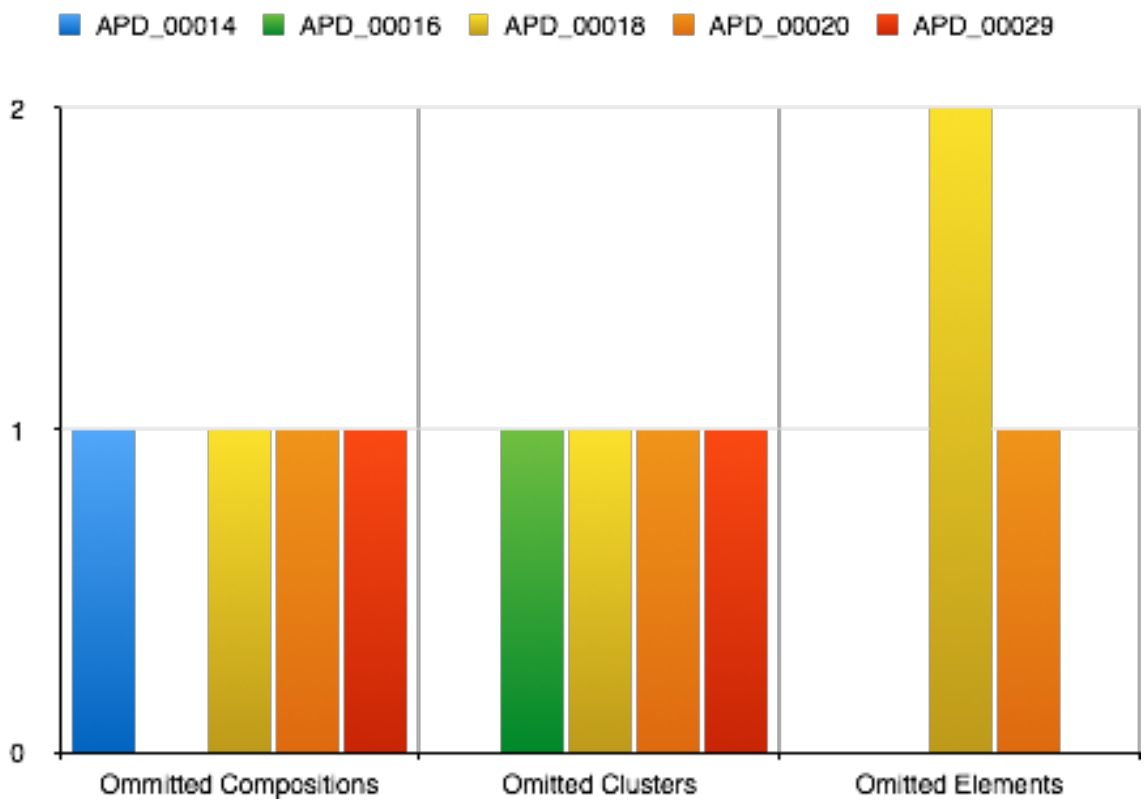
Graph 62: Number of Details added for Question 6 Sheet 2



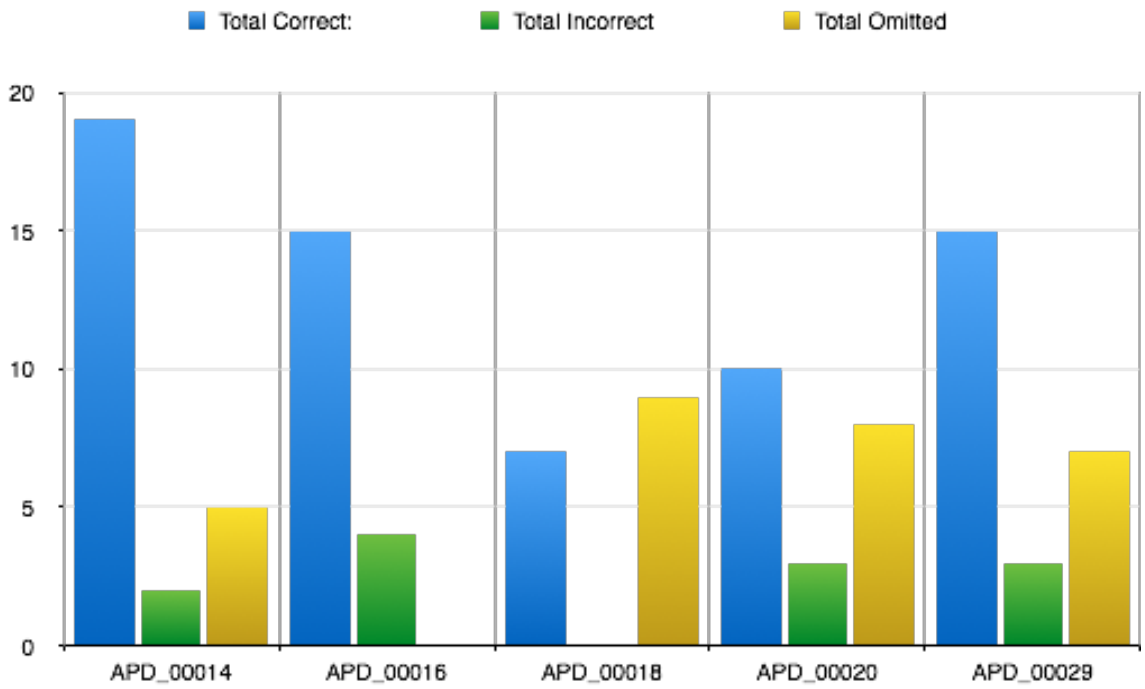
Graph 63: Number of correct 13606 classes added



Graph 64: Number of incorrect 13606 classes added



Graph 65: Number of omitted 13606 classes



Graph 66: Total Scores for Classes added in Question 6 Sheet 2

Graph 66 shows the total score for correct, incorrect and omitted Classes across all participants. There again appears to be higher proportion of omitted Classes to incorrectly specified ones. With the exception of participant APD_00018, all participants tended to score highest for correctly specified Elements.

Question 7 Exercise Sheet 2 - Security Breach

Chart 13 provides the breakdown of Understanding and misunderstanding for question 7 across all participants. In this case, there were as many cases of misunderstanding when authoring excerpts as there were for understanding, where there were ten in each case. There was one case of expectations being exceeded. Graph 67 again shows some variation in the numbers of Classes added by the participants, where the lowest was four and the highest was eighteen.

This question indicated a comparable degree of misunderstanding and understanding. Whilst there were a few errors in the Elements that were provided, these were essentially minor. The main reasons for misunderstanding were the omissions of Information Assets that provided details around the Information Assets of security codes and passes. It would remain to be seen if this

had any effect on participants answering the questions in Exercise 2. There was again some variation in the number of details added.

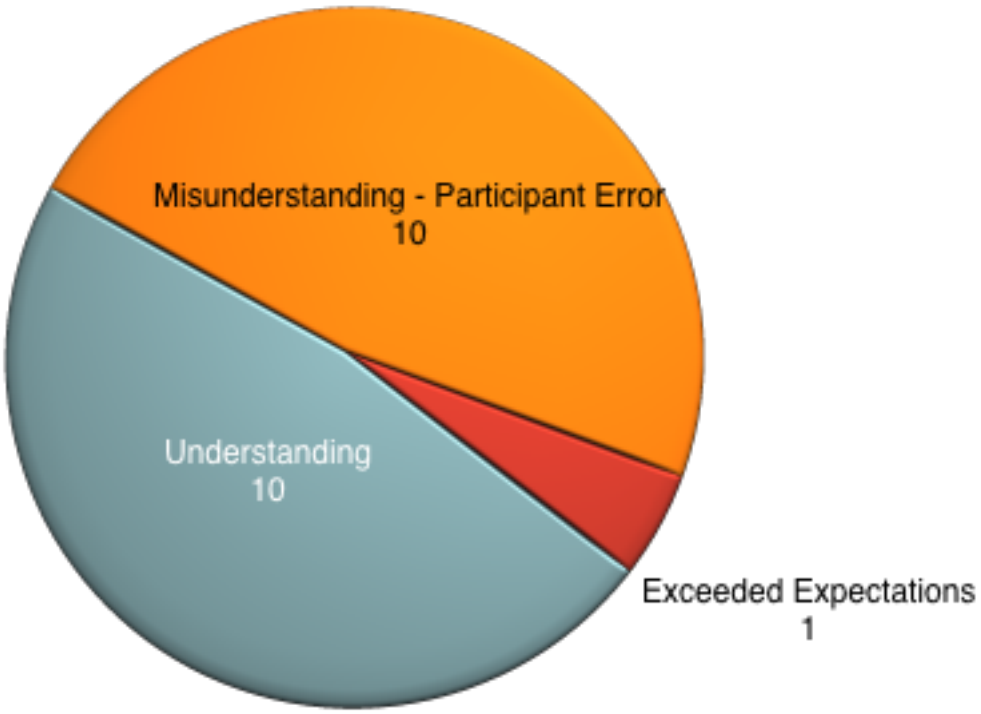
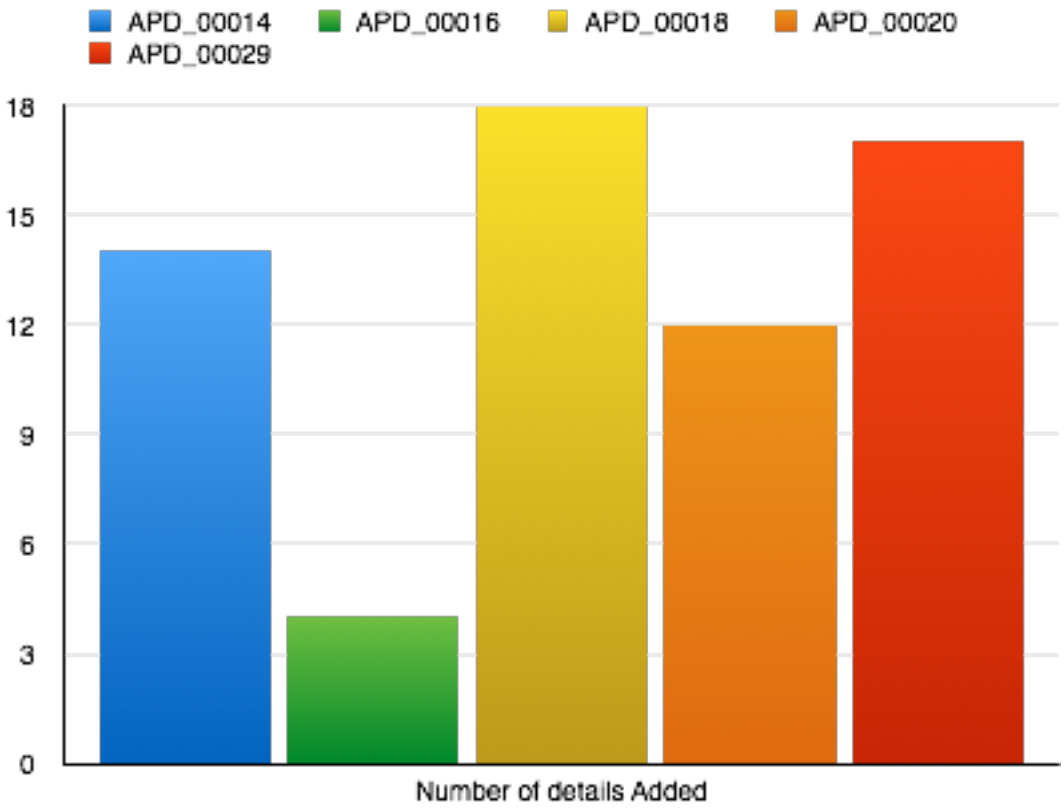
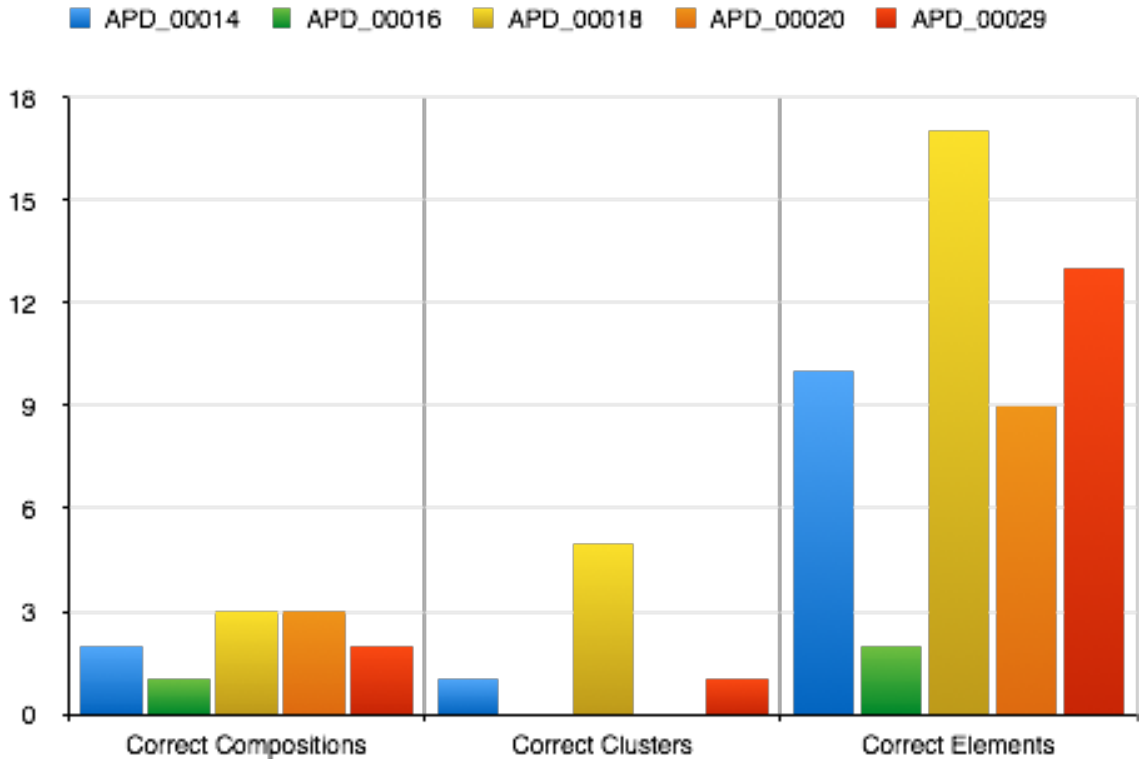


Chart 13: Measure of understanding for each participant, Question 7 Sheet 2

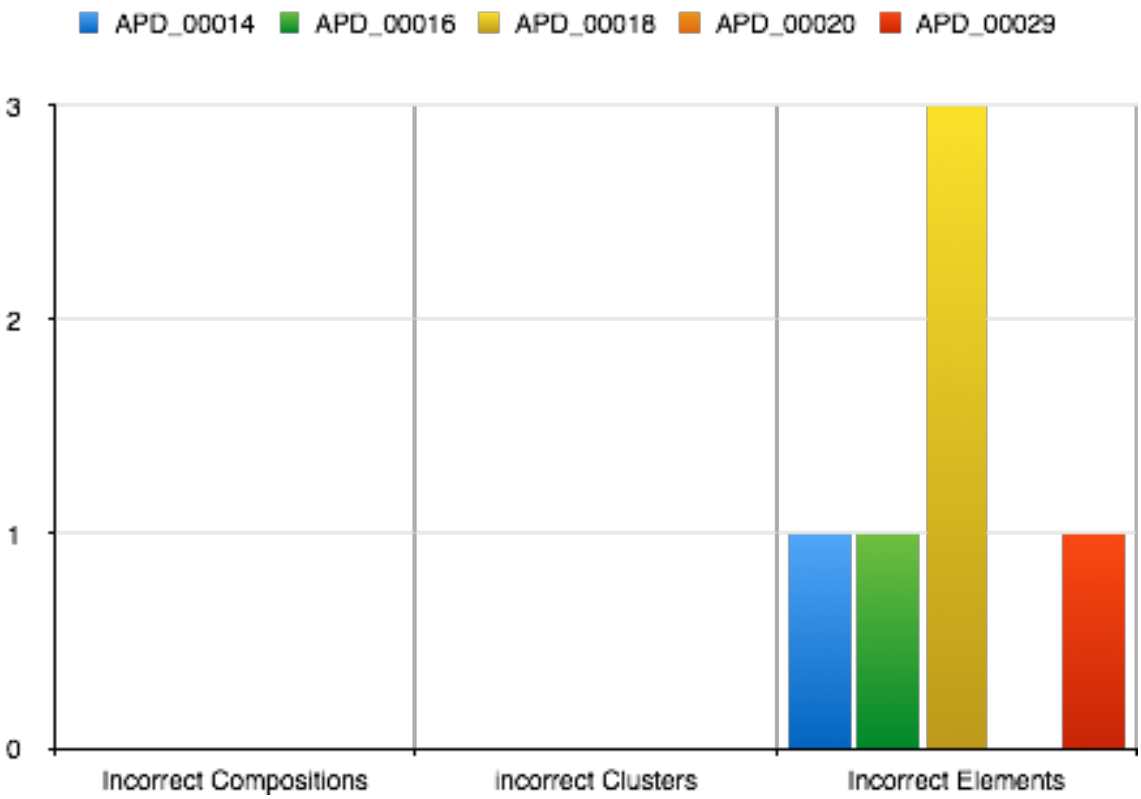


Graph 67: Number of Details added for Question 7 Sheet 2

Graph 68shows the number of correct Classes added for Question 7. There is some variation across the numbers of correct compositions, Clusters and Elements added, with three Compositions added as the maximum and one as the minimum. There is a variable number of Control Clusters, with a minimum of none and, in one case, three. There was a similar variation in the Elements, with a minimum of one and maximum of seventeen.

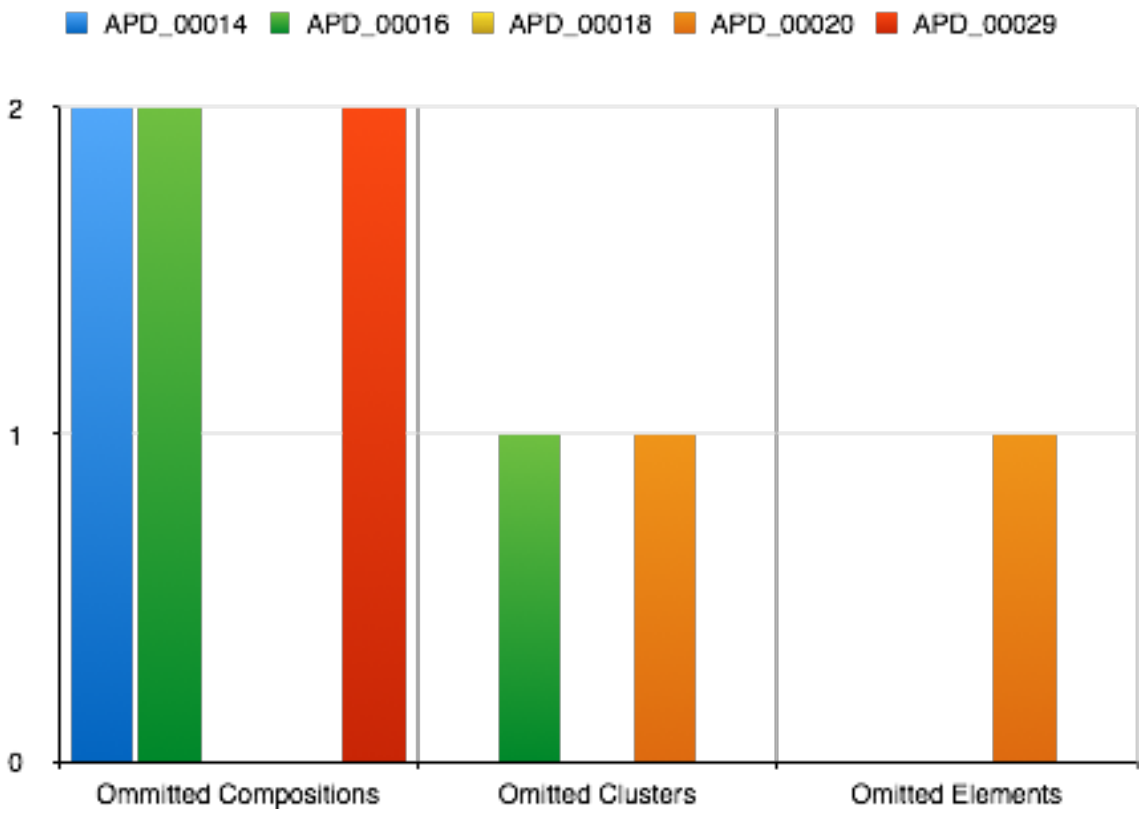


Graph 68: Number of correct 13606 classes added



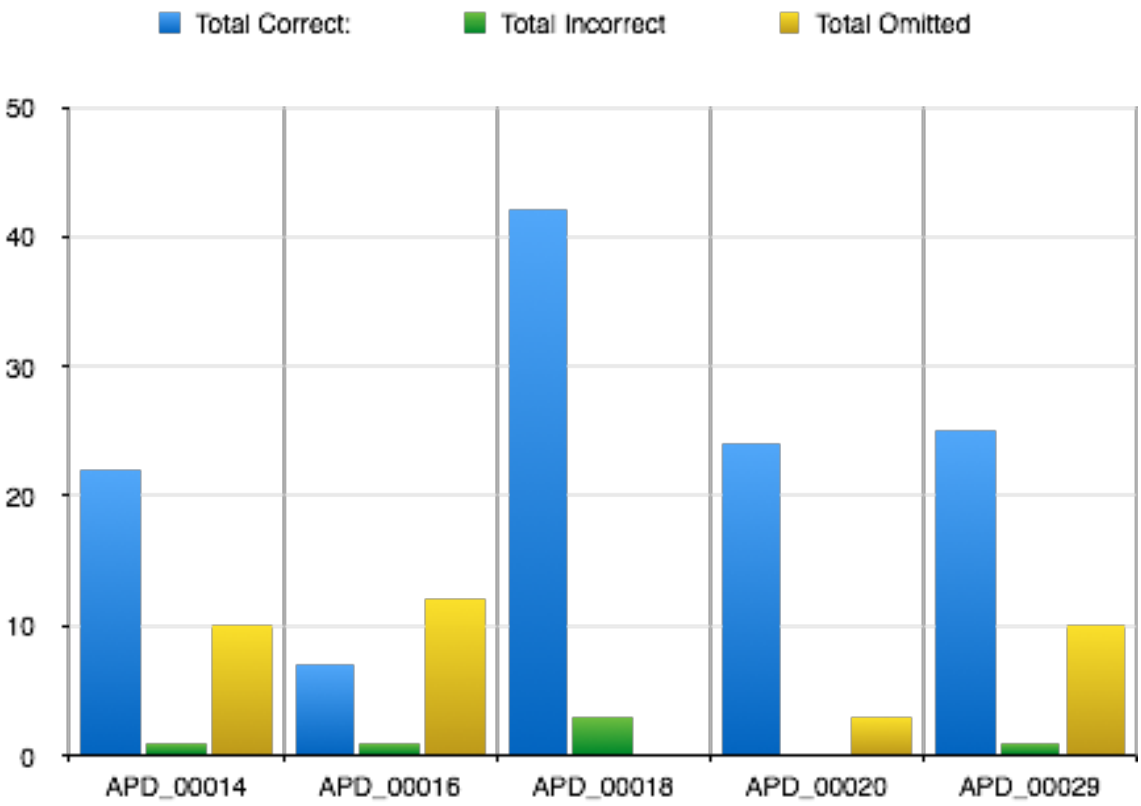
Graph 69: Number of incorrect 13606 classes added

Graph 69 provides the number of incorrect Classes added, in this case Elements only. The maximum of three incorrect elements appears is three, though this correlates to the highest number of added Classes from Participant APD_00018, which is a different to the lower number of Classes added by this participant in several of the earlier questions.



Graph 70: Number of omitted 13606 classes for Question 7, Sheet 2.

Graph 70 shows the number of omitted Classes across all the participants who tackled this exercise sheet. Three participants omitted two Compositions each, whilst two omitted one Cluster each. There was at least one omitted Element. As shown by Graph 71 below, there was a higher level of omitted Classes than incorrect ones, which is consistent with the results from the other questions. The highest total score for correct responses was forty-two, the lowest twenty-three. For errors, there were cases of a zero score as the lowest and three was the highest. For omissions, twelve was the highest and two the lowest. All participants made at least one omission in responding to this question, where in one case the omission score was higher than the correct score.



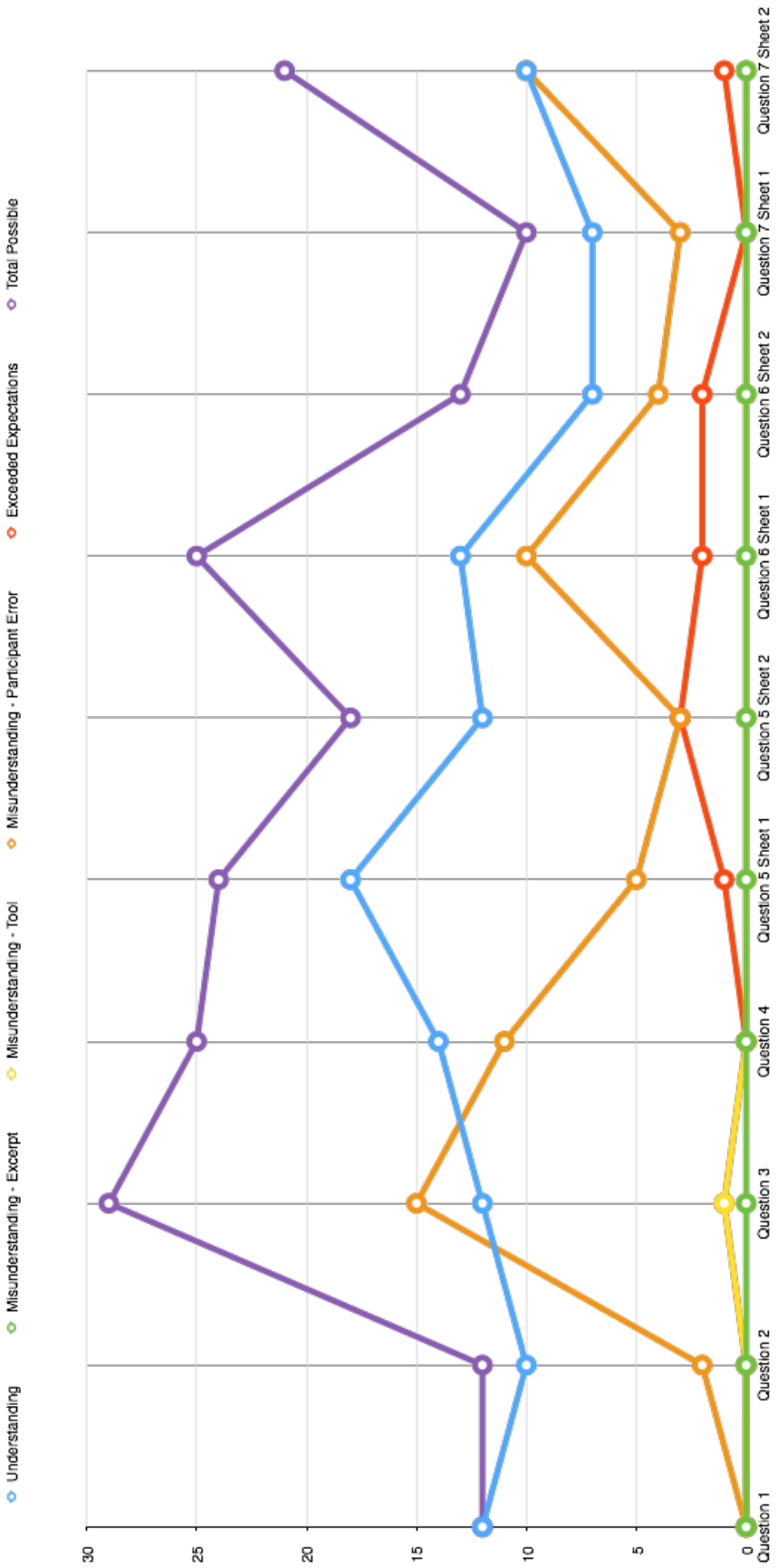
Graph 71: Total scores for correct, incorrect and omitted Classes, Question 7 Sheet 2

Overall Understanding Results Across Participants for Exercise 1

Graph 72 shows the overall scores for understanding across each of the questions. With the exception of question three, Participants tended to understand how to answer the questions more than they misunderstood them. The misunderstanding tended to be due to the participants making errors, and could be attributed to the varying complexity of the questions, learning to use *keibi* whilst attempting the questions, or misinterpreting the questions and how they related to the tool. This would not be unexpected given that very limited training was given prior to commencing the experiments as the author intended. Question three was intended as a warmup question, and it is possible that the higher misunderstanding score was due to participants getting used to the feature of adding multiple classes.

There was only one case where *keibi* itself introduced a misunderstanding. There was no apparent misunderstanding of the policy excerpts themselves. Participants also tended to exceed expectations from the fifth question, suggesting that they were starting to become more familiar with the tool, confident in its use

and that of the information model that they were presented with. These points are corroborated by some of the feedback from the third experiment.



Graph 72: Overall Results for Understanding Across All Questions

Graph 73 shows the overall numbers of understanding, misunderstanding and exceeding of expectations across all participants. This seems to follow a consistent trend, where participants generally appeared to understand the use of *keibi* to author policy, and where they misunderstood it was due to their error in interpreting and specifying the policy excerpts and use of the tool. This misunderstanding was less than the understanding demonstrated overall, and in some cases participants exceeded expectations in interpreting and authoring the excerpts.

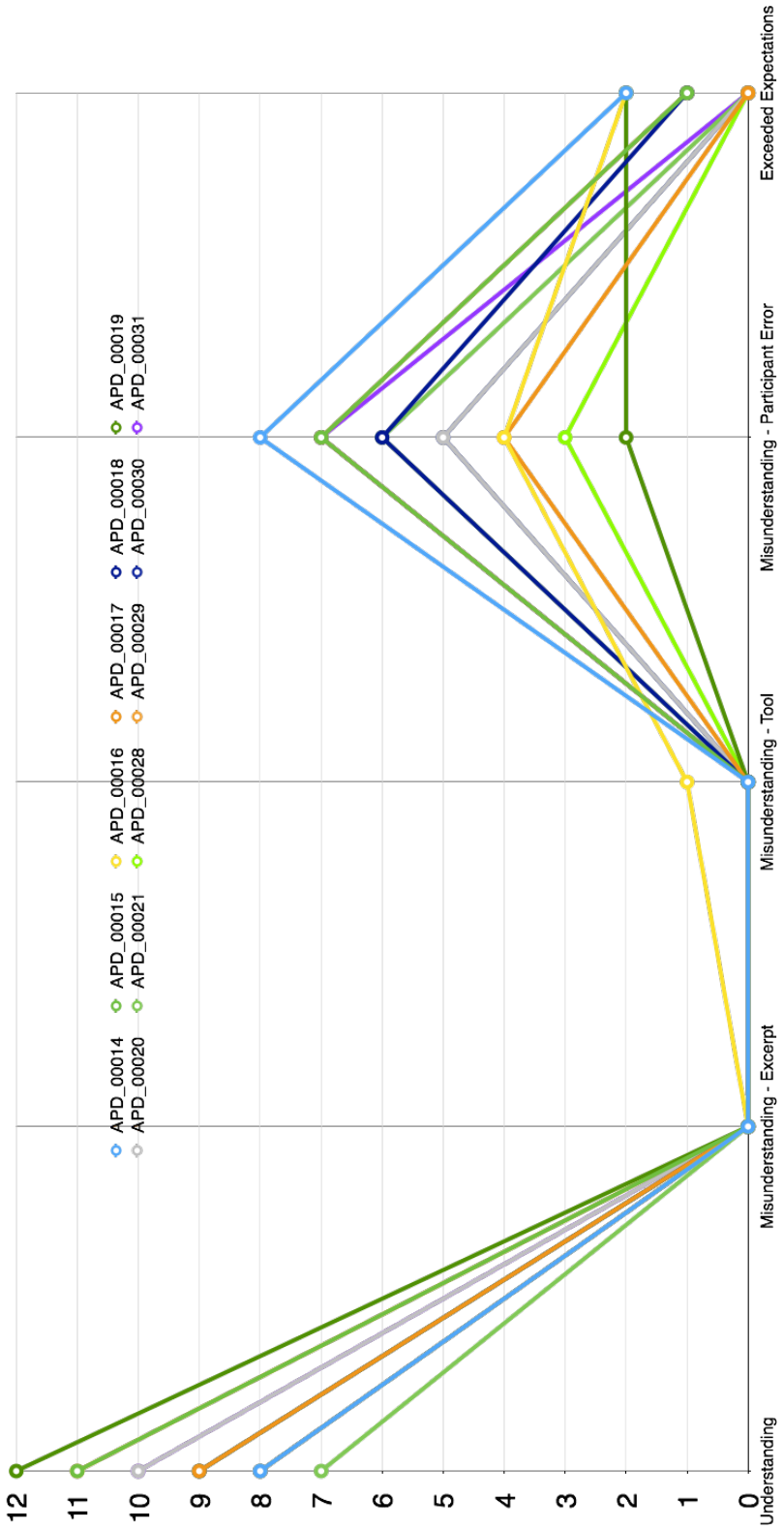
Summary of Results from Experiment 1

In all cases, participants opted to add the original excerpt or a summary of it in the Description field of the Safeguard (or Activity, if they interpreted the excerpt as an Activity only); this happened without any specific guidance to that effect. Participants tended to omit the same items (like Activities and Information Assets), though this was not always the case: Safeguards were also sometimes omitted. It was also notable that participants frequently omitted Controls from Safeguards, opting to allow the Description field to hold a summary or entire narrative of the original excerpt. This was partially explained by participants understanding the Safeguard to be the unit of Control in *keibi*. This suggests that additional guidance on how to use the different Compositions effectively would be useful.

This would show a variation of results for the second experiment, where there tended to be a variation in the numbers of details that the participants added: whilst this was not specific to a particular skill set or set of experiences, it did seem to depend on how many Elements in a given Composition were available. This shows that the participants had varying degrees of confidence about using the system and took either a cautious approach, entering what was necessary and sufficient, or chose to add further detail, which (aside from not always being correct) was neither expected or required.

There was a noticeable trend for participants to avoid adding Controls in the Safeguards. This was partially explained by participants understanding the

Safeguard to be the unit of Control in *keibi*. This suggests that additional guidance on how to use the different Compositions effectively would be useful.



Graph 73: Overall Understanding, Misunderstanding and Exceeding of Expectations Across Participants.

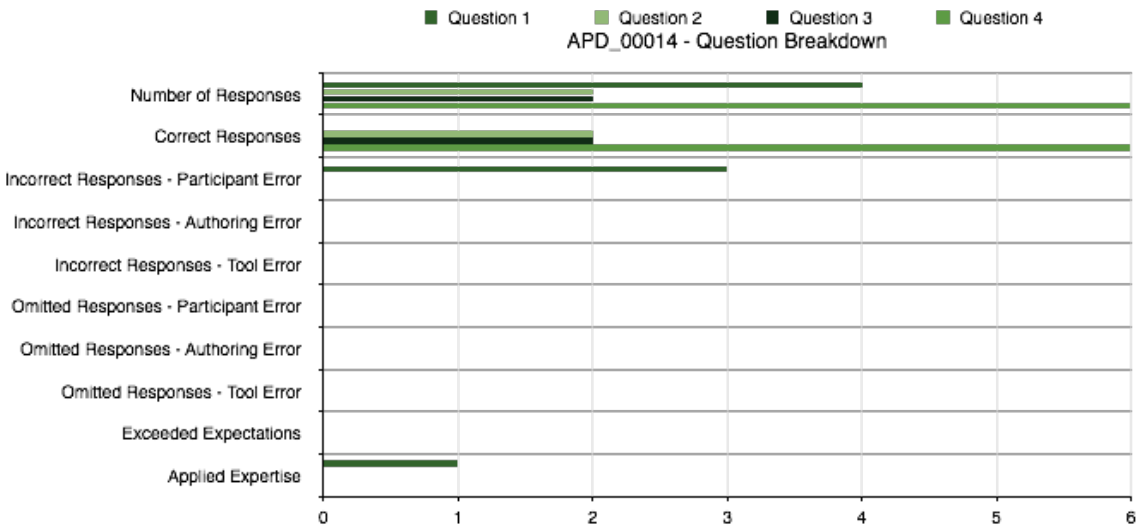
Appendix 19. Results gathered from Experiment Two

Results from Experiment Two

Each participant was asked to answer four questions about how they would respond to a given situation when consulting the details in *keibi*. The results would score the number of responses that they made and whether they were correct, incorrect, included additional detail based on their own abilities and / or whether they exceeded expectations. This section reviews the results for each participant in turn so that the outcomes can be reconciled with the excerpts that were used to answer each question: in each case, one participant used excerpts authored by their counterparts in each evaluation section, and the results of the authored experiment are compared to the results gathered in experiment 2.

Participant APD_00014

Graph 74 shows the breakdown of responses provided by participant APD_00014. The participant used the excerpts as authored by Participant APD_00019. They had largest number of responses, but only made an error in answering question one. Whilst there were some errors and an omission in some of the details as authored by APD_00019 in questions three and four, these related to errors in providing a computable set of specifications. There was nevertheless sufficient detail to answer the questions as expected. Participant APD_00014 claimed that they felt "a little lost" later in the group discussion, and felt that they had not quite appreciated where to look initially. This related to the steep learning curve that some participants had experienced when they first started to use the tool.



Graph 74: Number and breakdown of responses for participant APD_00014

Chart 14 shows the proportions of how participant APD_00014 responded across all questions. The majority of the responses ere correct, with some error as described. There was some evidence of the participant using their own expertise.

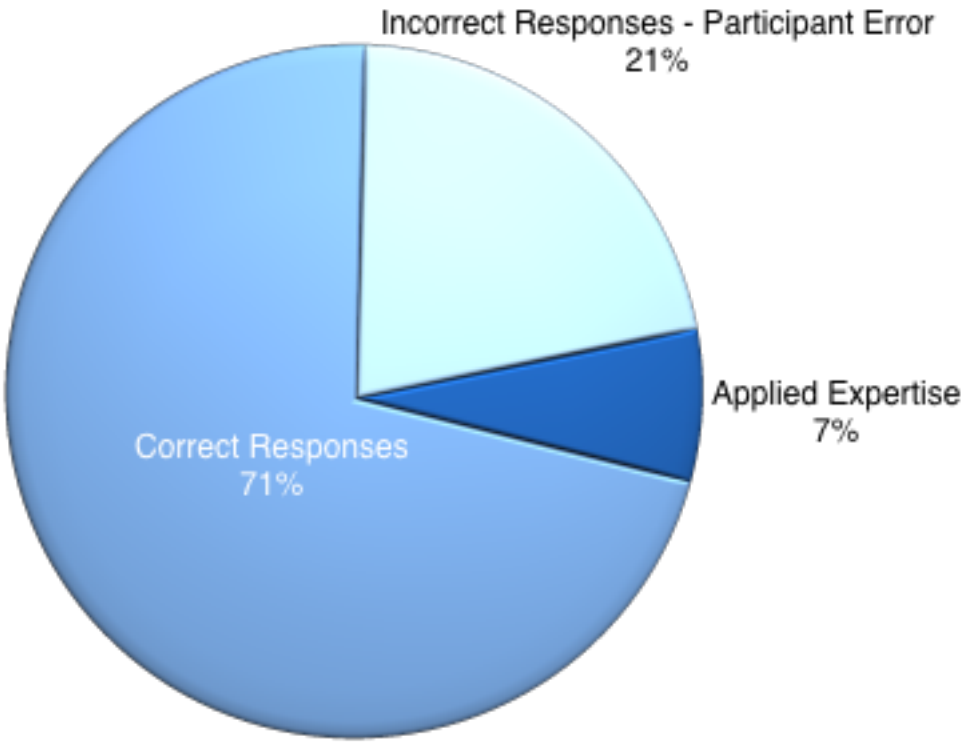
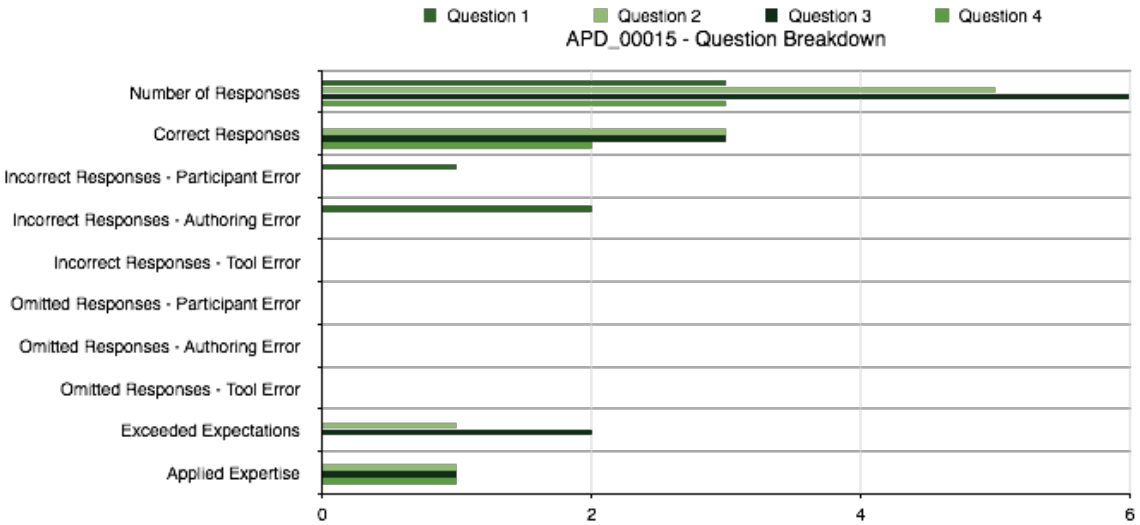


Chart 14: Breakdown of participant APD_00014's results across all questions

Participant APD_00015

Graph 75 shows the results of participant APD_00015's responses. APD_00015 correctly answered most questions, surpassing expectations and applying their own expertise in questions two, three and four. They made three errors in answering question one, however, and these were as a result of not seeing the details that had been placed in the description fields as provided by APD_00018. This is significant as APD_00018 had repeatedly added details to the description field only, not providing them in the other available Elements. This indicates that providing minimal details leads to those being missed by readers of the policy: whilst it is arguable that the reader here is at fault for missing the details buried in the narrative of the description, it should be noted that the *keibi* facilities are designed to make reading policy details easier to find and read. The number of responses tended to be lower than other participants - this is due to the minimal set added by APD_00018 and the ease of teasing out the details from the Description fields where most of the details were added.



Graph 75: Number and breakdown of responses for participant APD_00015

Chart 1 shows a breakdown of participant APD_00015's responses across all questions. This showed a greater variety of results, with just under a fifth of errors and about 36% of responses either relying on the participant's own expertise or their exceeding expectations. This was due to their having to rely more on their

own understanding of what was needed since the details entered on *keibi* were less detailed.

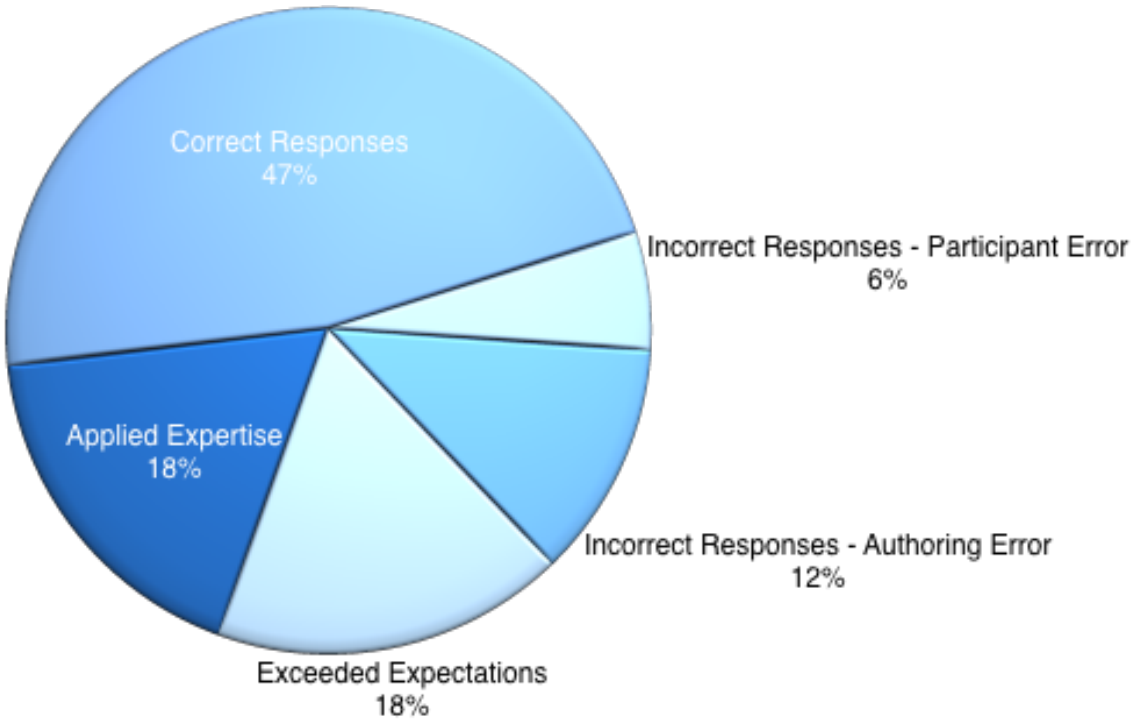
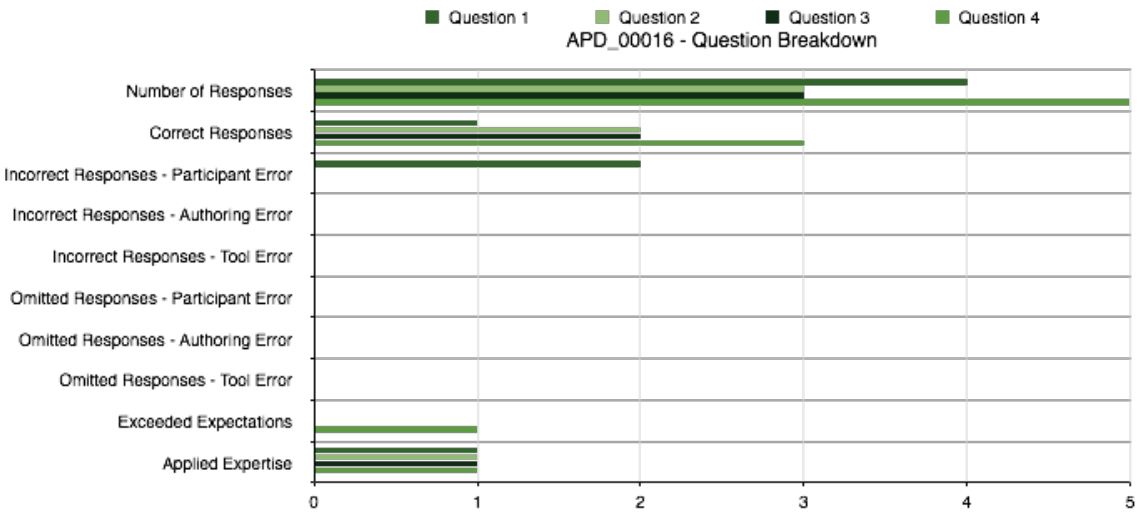


Chart 15: Breakdown of participant APD_00015's results across all questions

Participant APD_00016

Graph 76 shows the results of participant APD_00016's responses. APD_00016 answered all questions correctly and applied their own expertise in each case. Although there are some omissions in the policies authored by APD_00017, the participant was able to find the details they needed to answer the question and demonstrate a good understanding of what was required by applying their own expertise and surpassing expectations. The number of responses tended to be higher than other participants, supported by the additional expertise applied by the participant.



Graph 76:Number and breakdown of responses for participant APD_00016

Chart 16 below shows overall a majority of correct answers and over a quarter of the responses were based upon applying their own expertise. Whilst this is a higher proportion than participant APD_00015, the number of omissions in the authored policies authored by APD_00017 than those added by APD_00018, suggesting that in this case, *keibi* helped to encourage APD_00016 to use their own expertise.

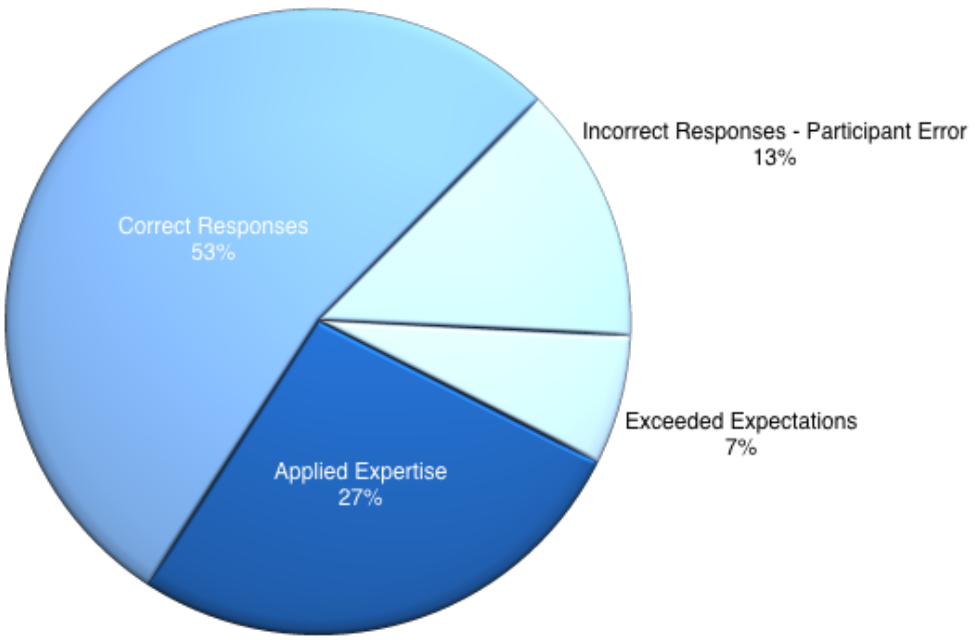
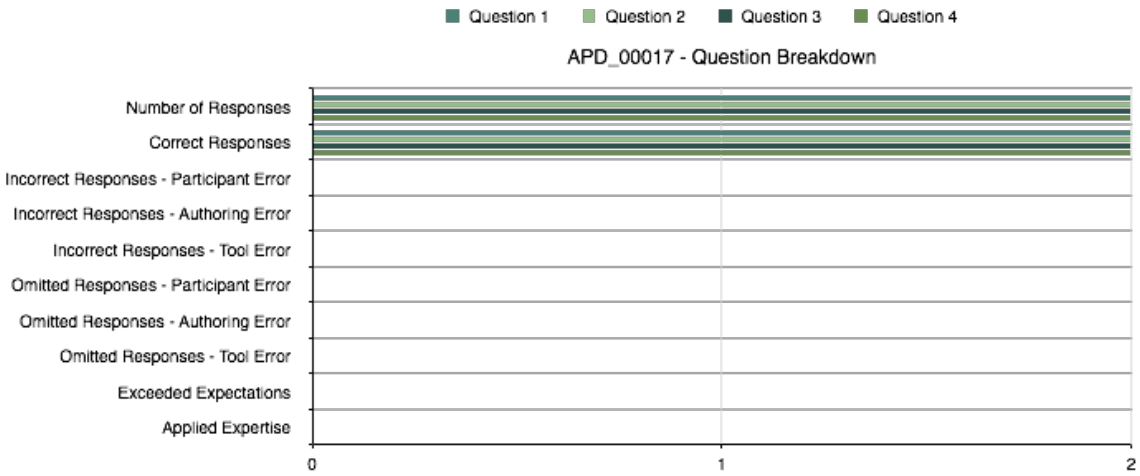


Chart 16: Breakdown of participant APD_00016's results across all questions

Participant APD_00017

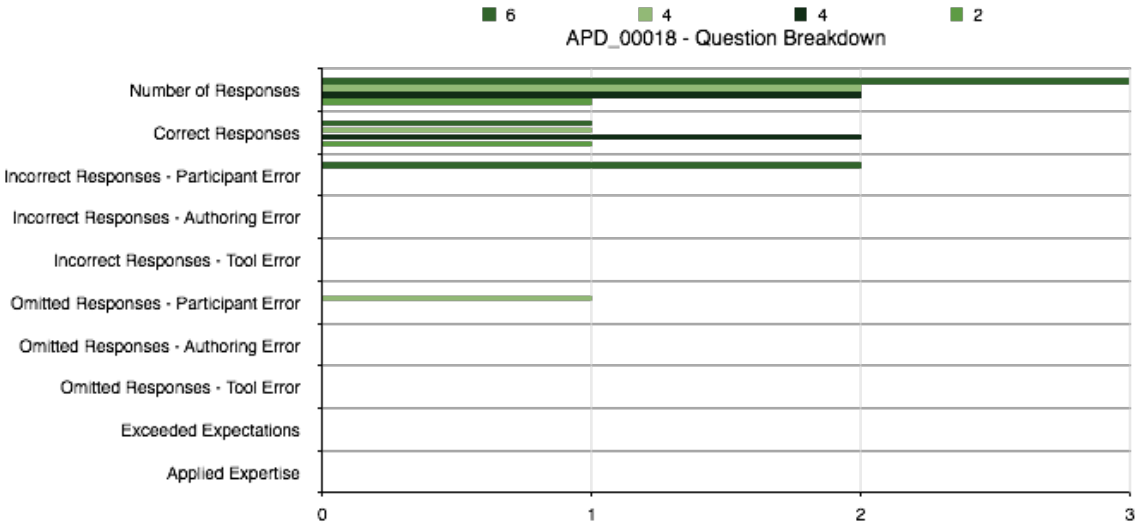
Graph 77 shows the participant was able to answer all questions with correct responses, but had the fewest responses. This was due to the limited number of details added by APD_00021 after their unsatisfactory experience in authoring policy and losing an entire Safeguard when they navigated to a help screen having not saved the details. APD_00017 was able to answer correctly, but without the detail of the other participants. This shows that *keibi* currently allows an under specification of policy. This is a point that was picked up in the group discussions where participants identified that a wizard or guidance on how to write a policy would be a very useful and powerful feature.



Graph 77: Number and breakdown of responses for Participant APD 00017

Participant APD_00018

APD_00018 generally answered questions correctly as shown in Graph 78, but made a lower number of responses and an incorrect response, as well as an omission. These were as a result of their error, and this can be explained by their concern raised in the satisfaction questionnaire as well as during the group discussion that they were not confident that they were using the tool correctly and had some trepidation. This suggested that some reassurance, guidance or other training might be particularly helpful in such cases.



Graph 78: Number and breakdown of responses for participant APD_00018

Chart 17 shows the breakdown of responses across all questions for participant APD_00018. The higher proportion of of errors is reflected, where the participant felt that they were not confident about using the tool correctly. This clearly affected their ability to apply their own expertise or surpass expectations, and the reasons for their lack of confidence are explored in the results for experiment three.

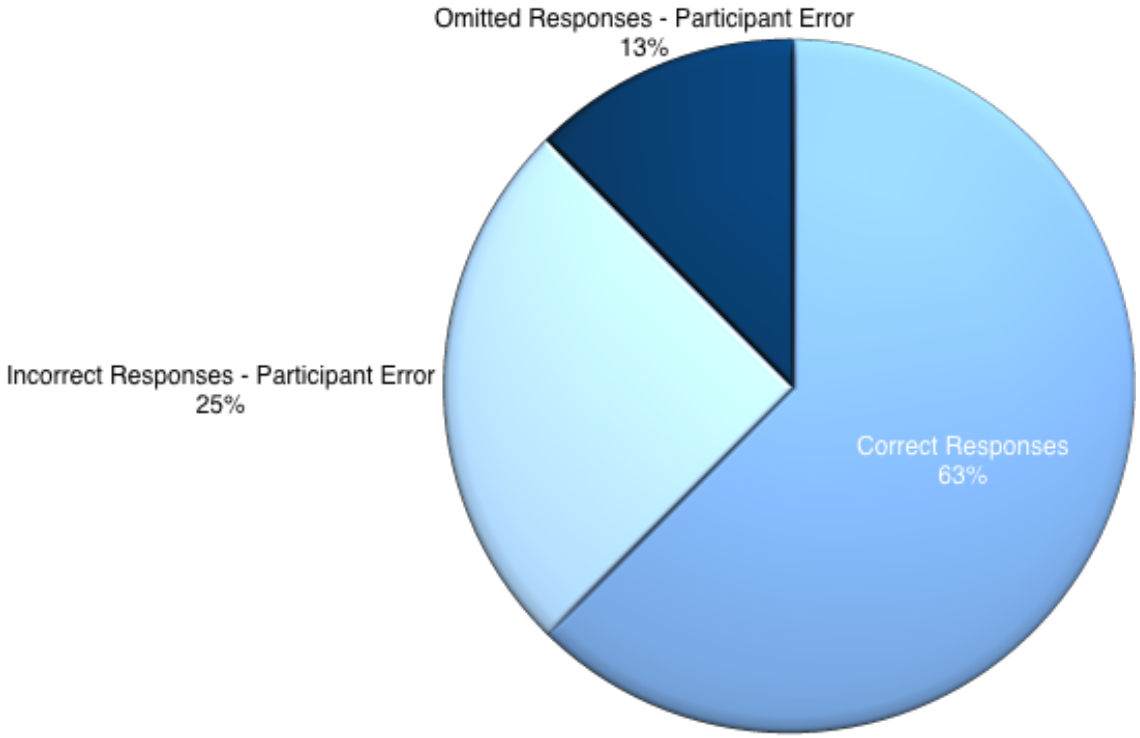
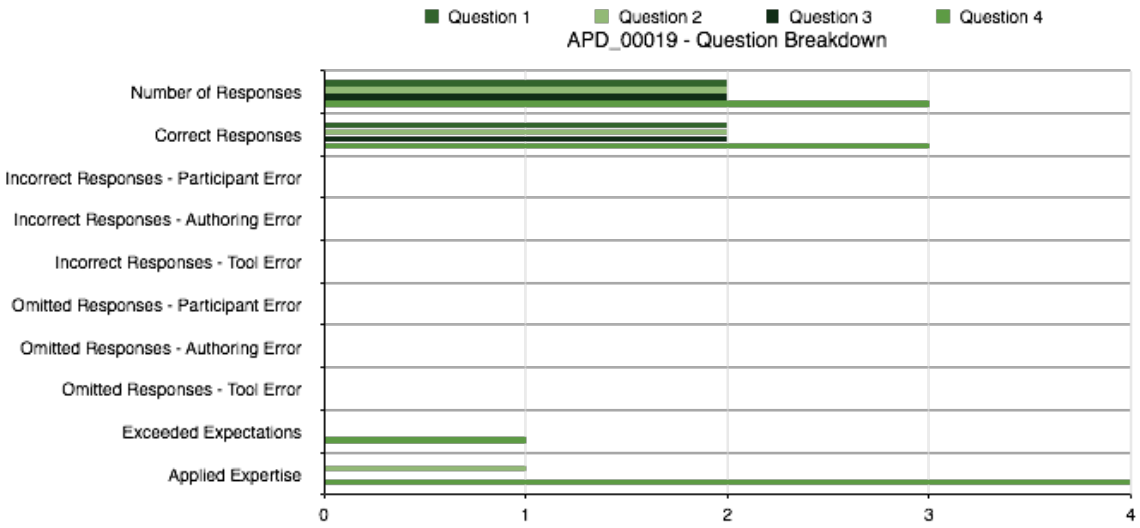


Chart 17: Breakdown of participant APD_00018's results across all questions

Participant APD_00019

Graph 79 shows the breakdown of responses for participant APD_00019 for each question. Whilst making a low number of responses, APD_00019 was able to apply their own expertise and surpass expectations using the details authored by APD_00014. The details that had been authored were unexpected and whilst there was some omission, APD_00019 was able to apply their expertise in a larger number of cases as shown in Chart 18. It is likely that this participant's experience as a data manager and service provider allowed them to apply their own expertise to answer the questions.



Graph 79: Number and breakdown of responses for participant APD_00019

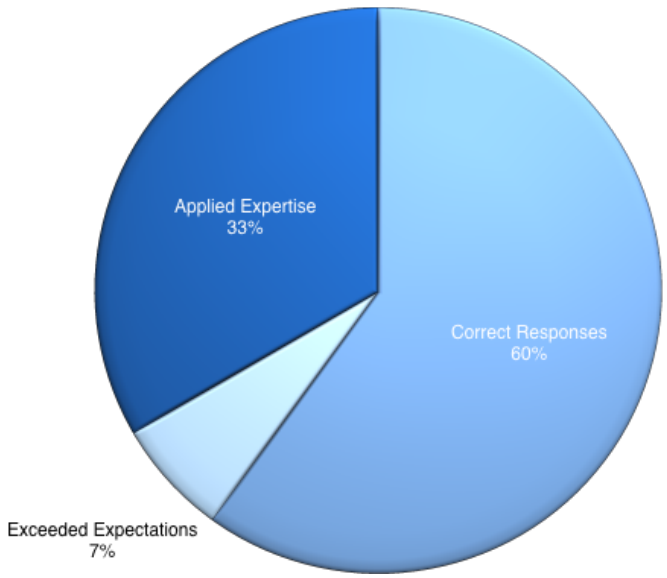
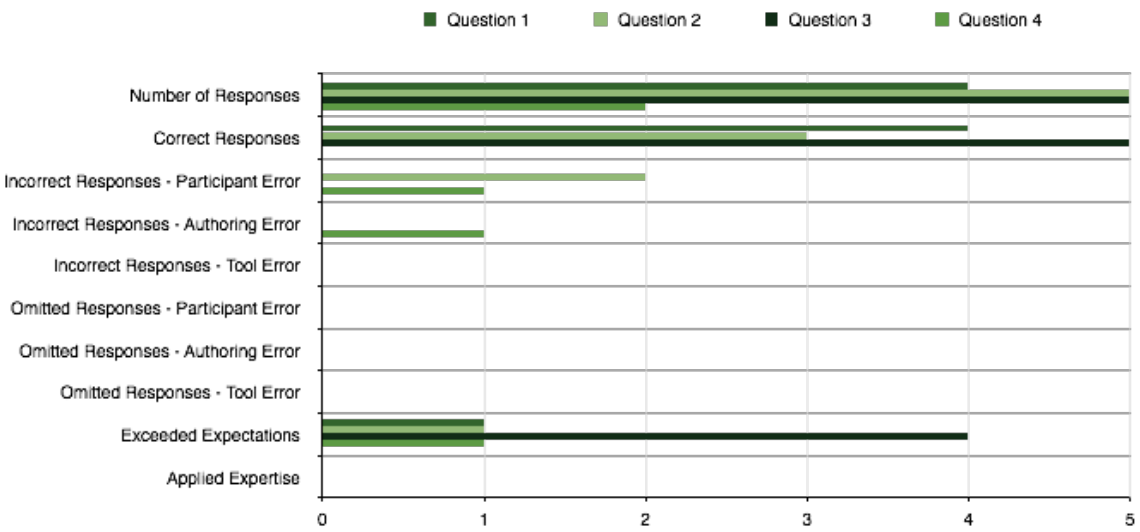


Chart 18: Breakdown of participant APD_00019's results across all questions

Participant APD_00020

Graph 80 shows the results for participant APD_00020 was able to provide a large number of responses, with only one error. They also exceeded expectations the most. It is possible that this was because of their experience as an IT professional for thirty years and the role that they occupy as a service manager. In addition to this, the participant used details authored by APD_00028, which had a significant number of Element and Control Omissions. Unlike APD_00015, however, the participant was able in this case to review many of the details within the Description fields and answer fully and correctly. This suggests that whilst some users will be able to see details in narrative text with more success than others, explicit definition should be provided. Chart 19 provides the breakdown of results for participant APD_00020. The proportion of error was nearly a fifth in this case, but there was a higher proportion of exceeding expectations based upon the use of *keibi*.



Graph 80: Number and breakdown of responses for participant APD_00020

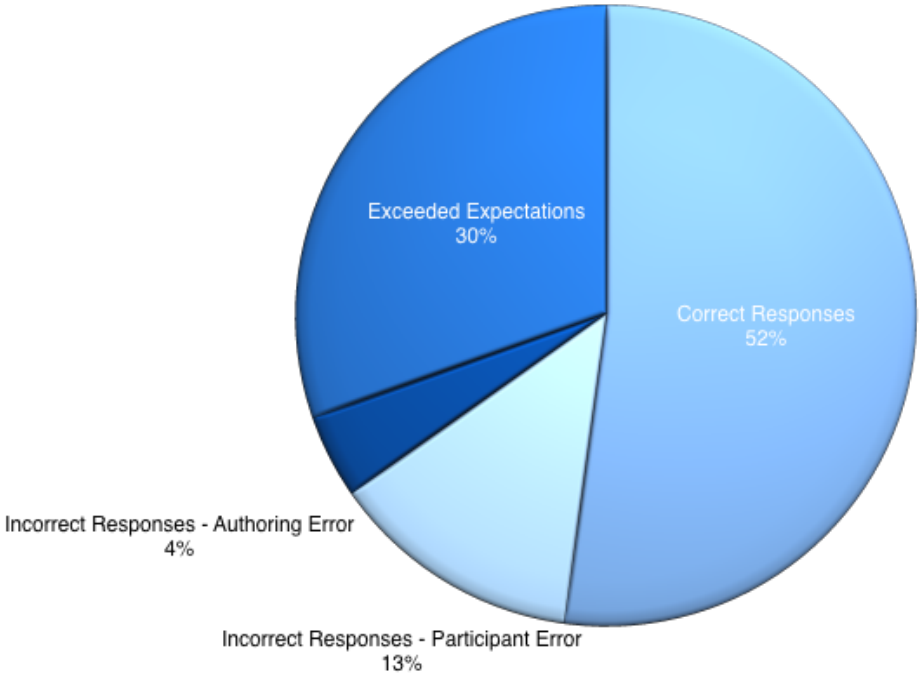
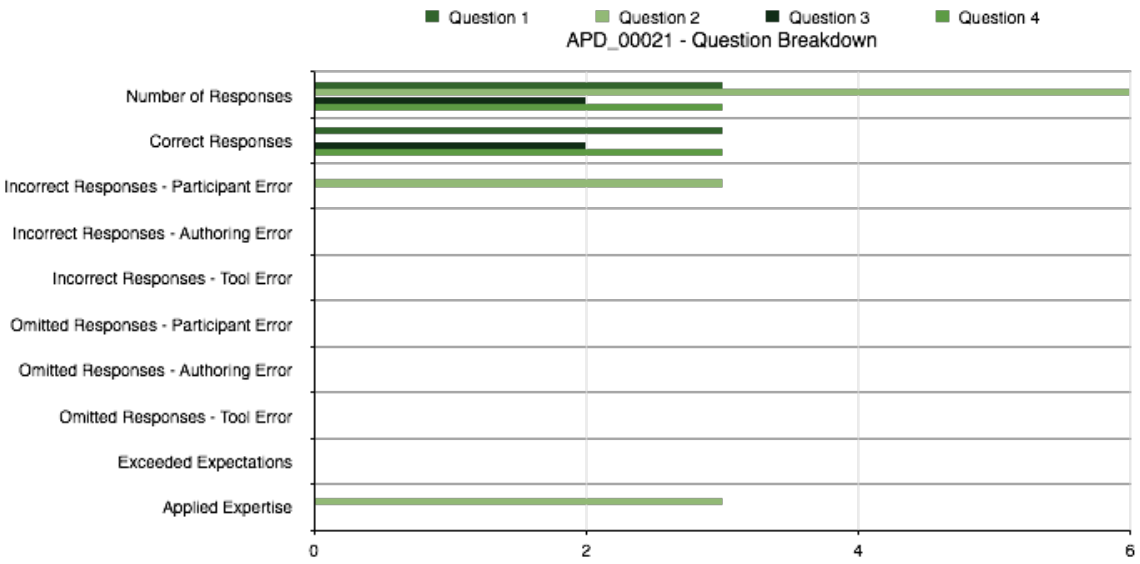


Chart 19: Breakdown of participant APD_00020's results across all questions

Participant APD_00021

Graph 81 shows the rsults for participant APD_00021. The participant was able to respond correctly and apply their own expertise to answering the questions. With the exception of answers to question two, the number of responses was lower, and this suggests that the participant had some difficulty in finding and applying the details in *keibi*. They also answered question two mostly using their own experience and did not refer to the details in *keibi*.



Graph 81: Number and breakdown of responses for participant APD_00021

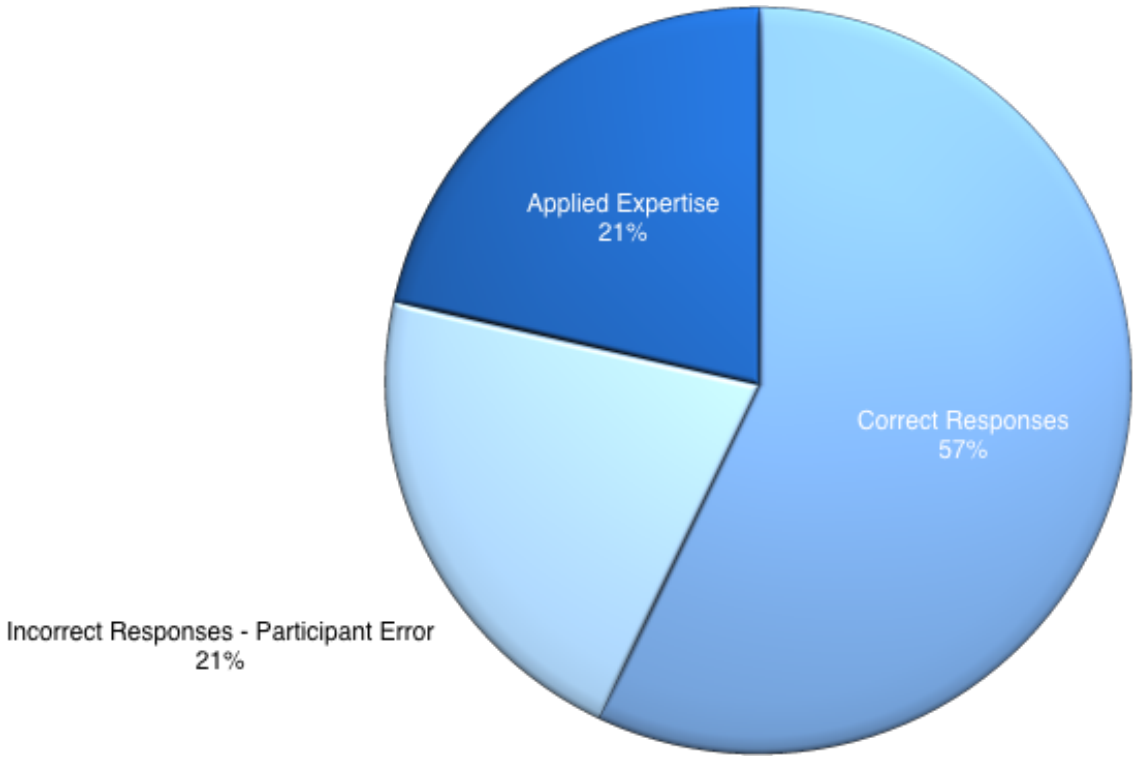
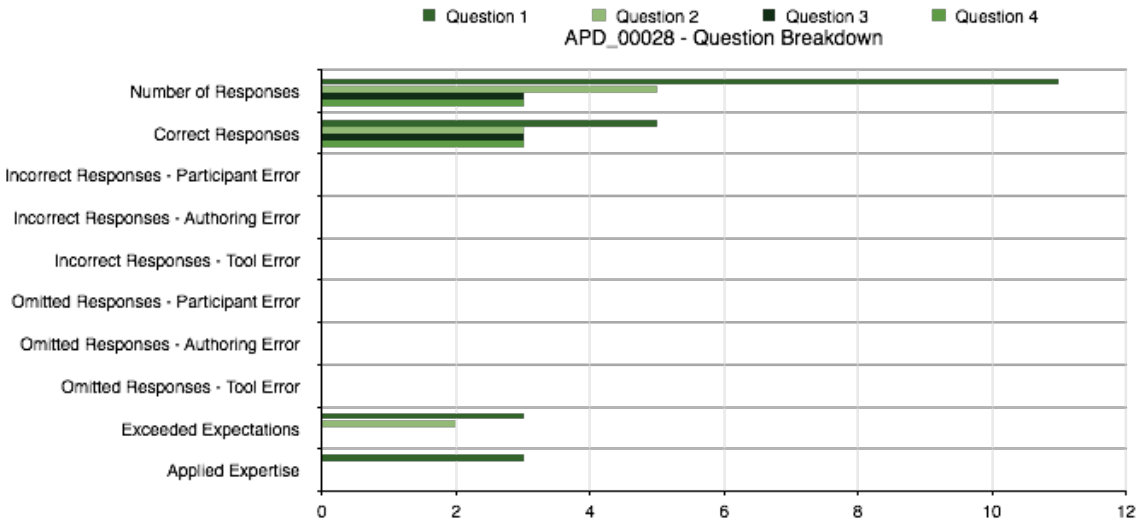


Chart 20: Breakdown of participant APD_00021's results across all questions

Chart 20 illustrates the higher level of incorrect responses, equal to the proportion of their applying their own expertise to answer.

Participant APD_00028

Graph 10 shows the breakdown of results for participant APD_00028. The participant answered the questions correctly and in some detail, applying their own expertise and exceeding expectations. Their experience suggests an explanation for exceeding expectations and applying their own expertise. They referred to policy items authored by APD_00030, which were less detailed and contained some omissions, but they were able to find the required details and answer effectively.



Graph 82: Number and breakdown of responses or participant APD_00028

Chart 21 shows the breakdown of results for participant APD_00028 across all questions. This illustrates that the majority of responses were correct and the participant was able to surpass expectations in almost a quarter of the responses, whilst also applying their own expertise to answer questions in almost a sixth of cases, particularly where the details entered into *keibi* had been omitted. This suggests that the use of *keibi* helps to support users with particular experience of data management and governance and security compliance to exceed expectations, though they must still rely on their own experience where details are not included.

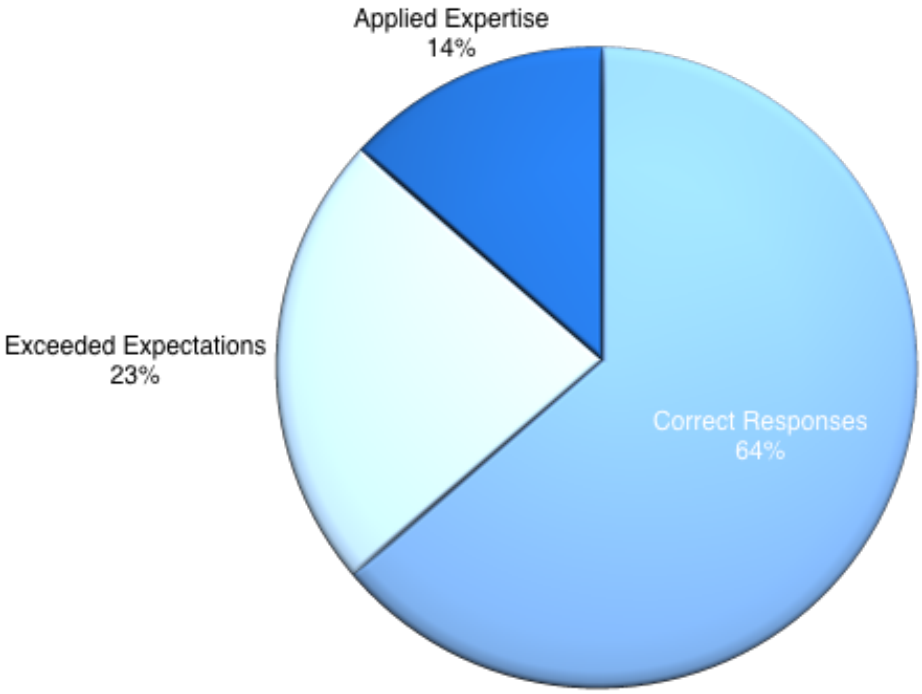
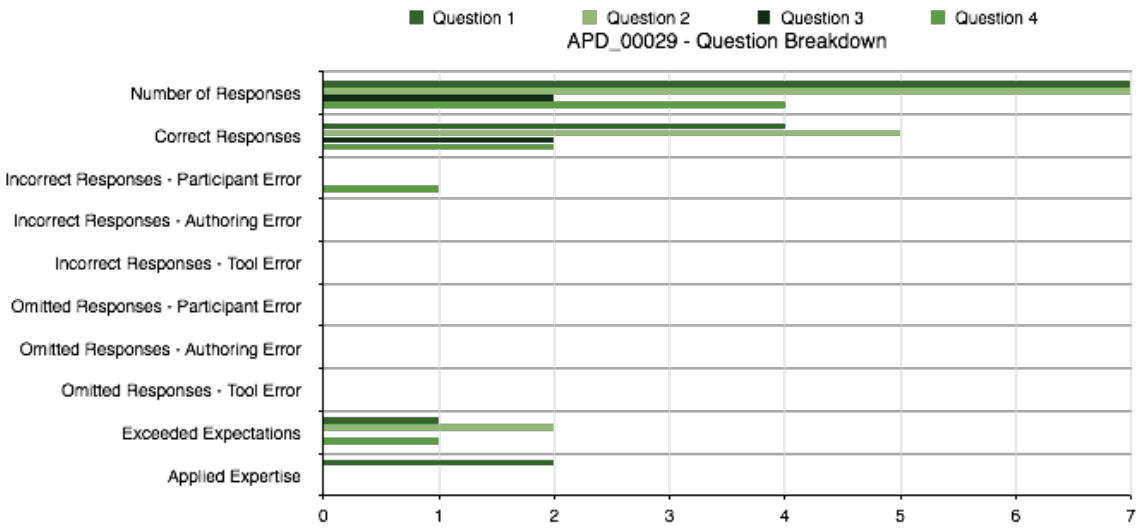


Chart 21: Breakdown of participant APD_00028's results across all questions

Participant APD_00029

The participant managed to answer all questions correctly with the exception of one, as shown in Graph 83. They also managed to exceed expectations and apply their own expertise in one case. The error is of interest - even though the excerpt authored by APD_00030 lacked some detail, this was not the cause of the error. The error was as a result of adding a proposed response, which breached a rule that had been specified in another Safeguard. This is reflected in Chart 22 suggests that the participant might have had trouble remembering details from across the different information sources. This point was picked up in one of the group discussion sections: there was some concern that it would become difficult to find the details that were needed should the number of items that are stored grow to a large number.



Graph 83: Number and breakdown of responses for participant APD_00029

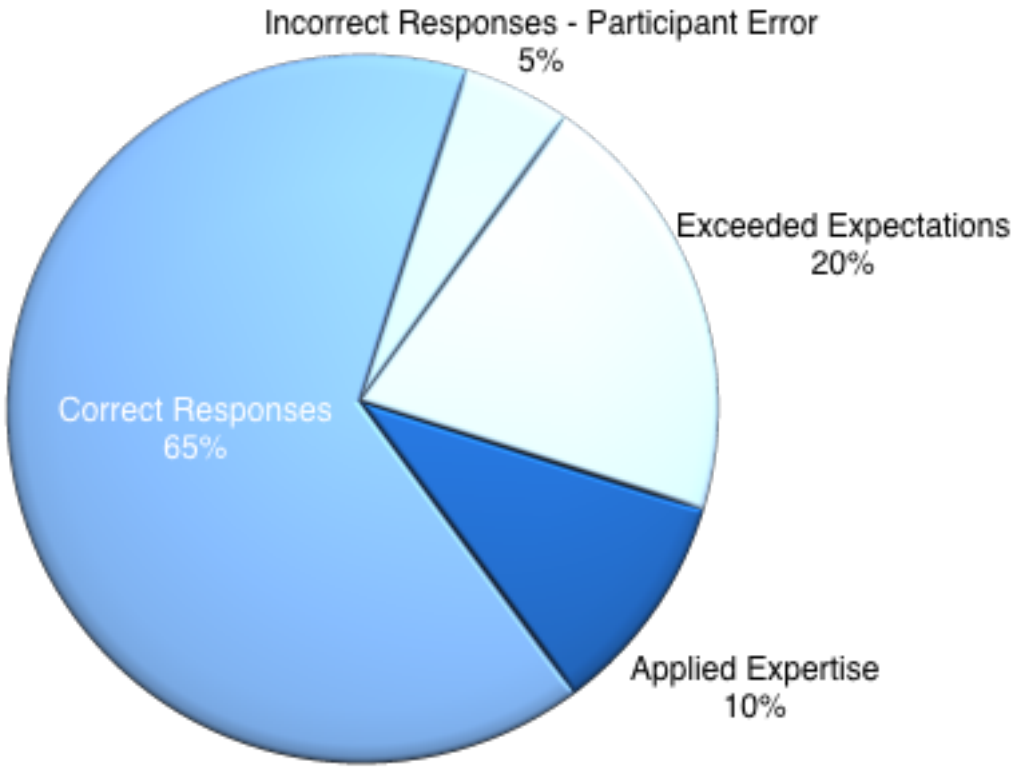


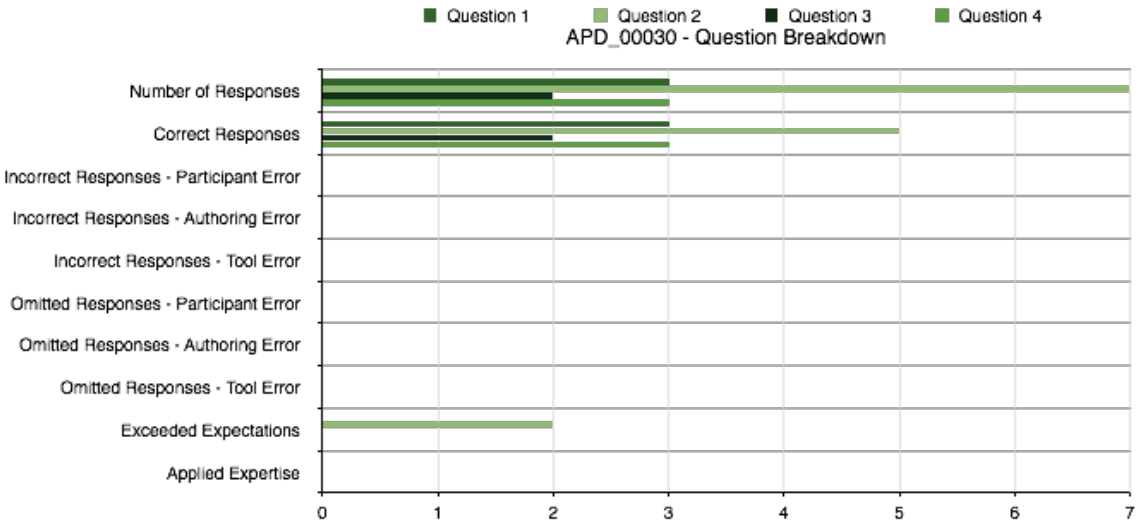
Chart 22: Breakdown of participant APD_00028's results across all questions

Participant APD_00030

Graph 84 shows a breakdown of the results for participant APD_00030, who had one of the higher numbers of responses and these were all correct, exceeding expectations twice. Whilst the policies authored by APD_00029 had some omissions and were in some cases incorrect, APD_00030 still managed to answer

the questions using the pertinent details. This suggests that there are some core items of information that participants will need in a given situation. This point was picked up in the Group Discussions - participants felt that having tailor made lists of policy items and expectations would be helpful, and that offering guidance on how to prepare the details would help to provide pertinent details.

Chart 23 provides the breakdown of results for participant APD_00030 across all questions. No errors were made, and the majority of responses were form using the tool in addition to the participant exceeding expectations.



Graph 84: Number and breakdown of responses for participant APD_00030

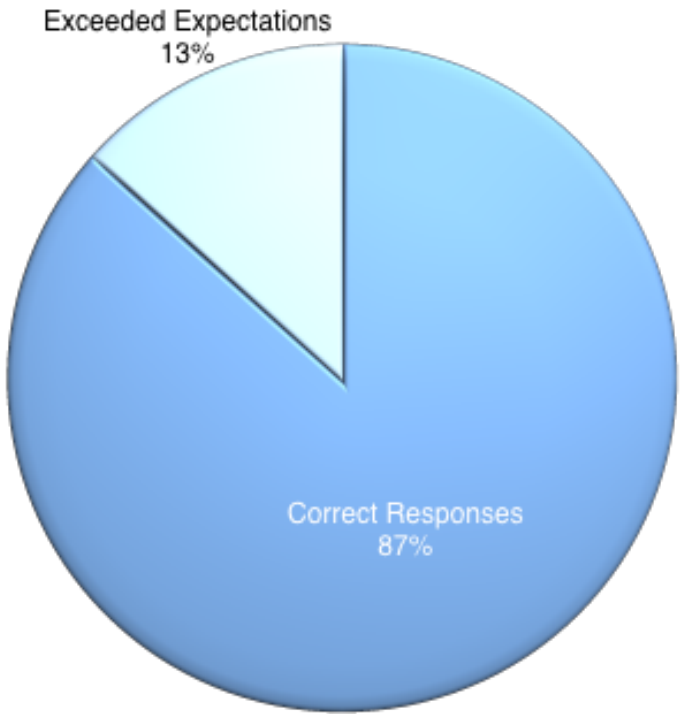
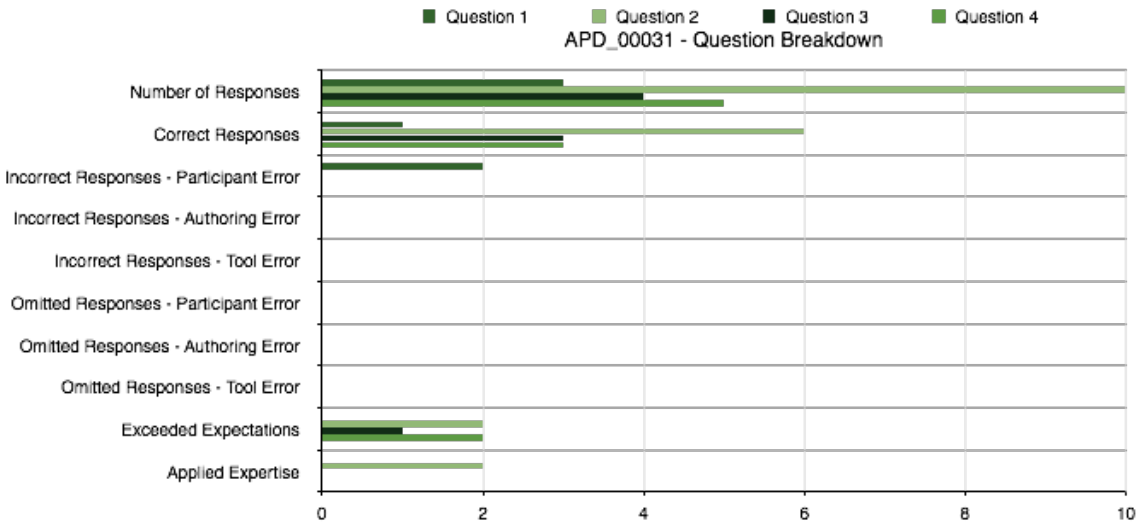


Chart 23: Breakdown of participant APD_00030's results across all questions

Participant APD_00031

Graph 85 shows that participant APD_00031 was able to provide a larger number of responses, exceeding expectations and providing correct responses, with the exception of two errors for question one. There was some evidence that this participant answered question using their own opinions rather than looking at the details in *keibi*, despite the fact that the answer was provided in the policy authored by APD_00020.



Graph 85: Number and breakdown of responses for participant APD_00031

Chart 1 APD_00031 applied their own expertise in just under a tenth of their responses, questioning the policy statement itself. As with other participants, nearly a fifth of their responses exceeded expectations. Though a tenth of the responses were incorrect, nearly sixty per cent were correct.

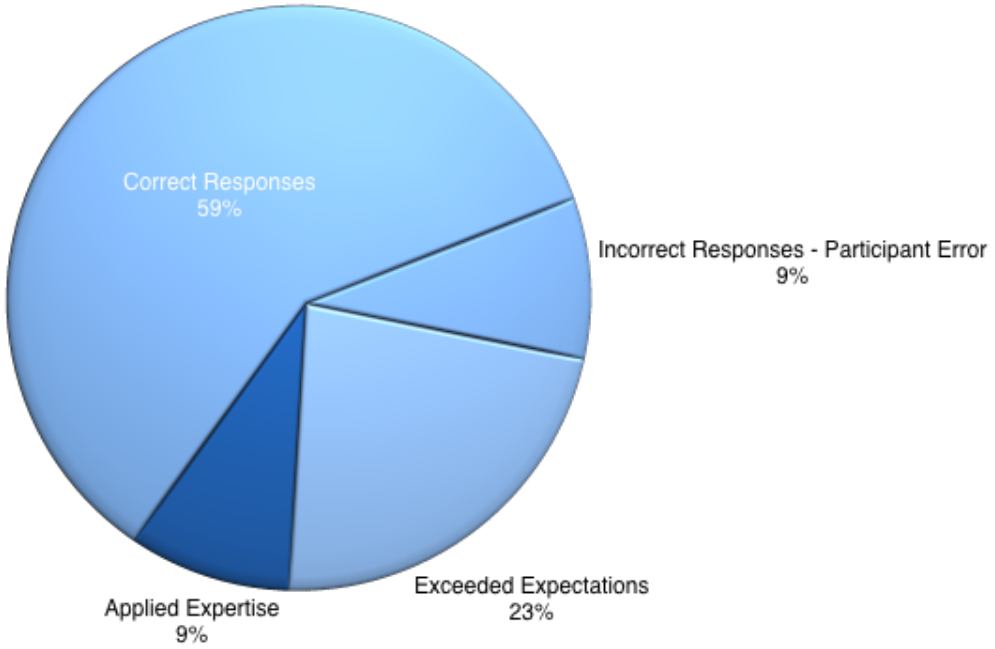


Chart 24: Breakdown of participant APD_00031's results across all questions

Overall Results for Each Question

The following four charts provide the proportions of results for each question respectively across all participants. This shows how participants fared overall, providing a combined view of their performance and an indication of how they responded to each question, be it to answer correctly, incorrectly, applied their own expertise or exceeded expectations.

Chart 25 shows the overall breakdown of responses. A total of fifteen per cent are errors, just over a quarter showed evidence of participants applying their own expertise or exceeding expectations and nearly sixty per cent being correct. Chart 26 shows the overall results breakdown for question 2, where there is a slightly higher proportion of exceeded expectations and application of their own expertise, a lower proportion of incorrect responses and the same proportion of correct answers as the first question. There was a small proportion of four per cent for omitted answers in this case.

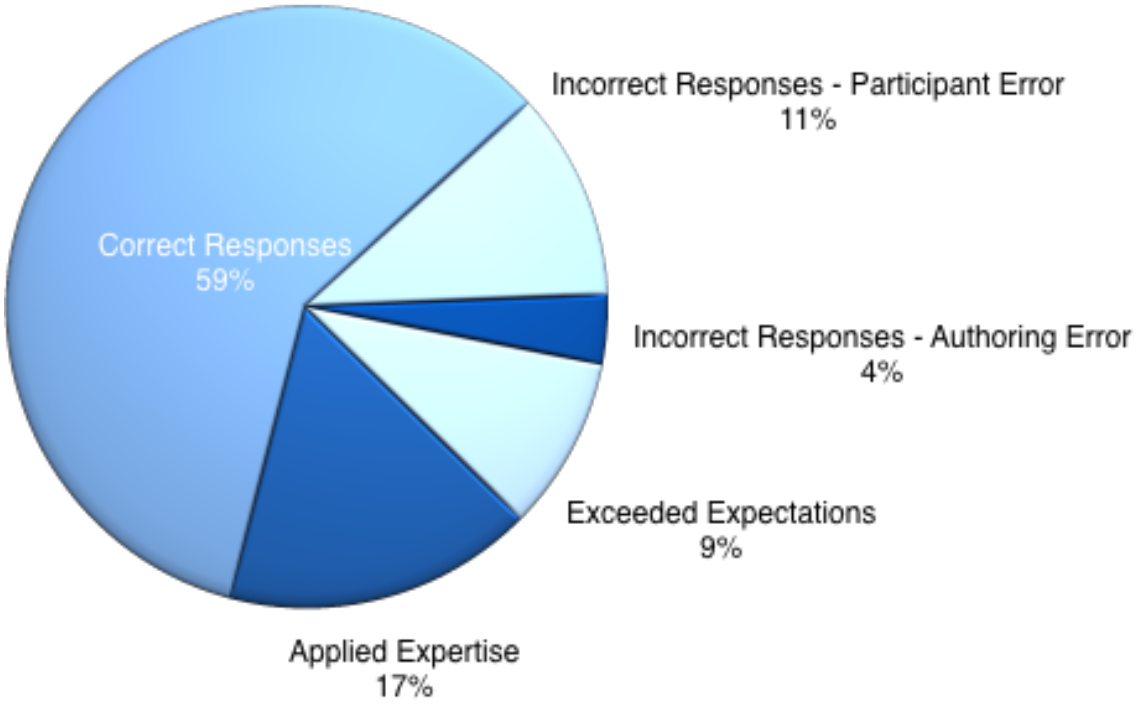


Chart 25: Overall breakdown of results for question 1 across all participants

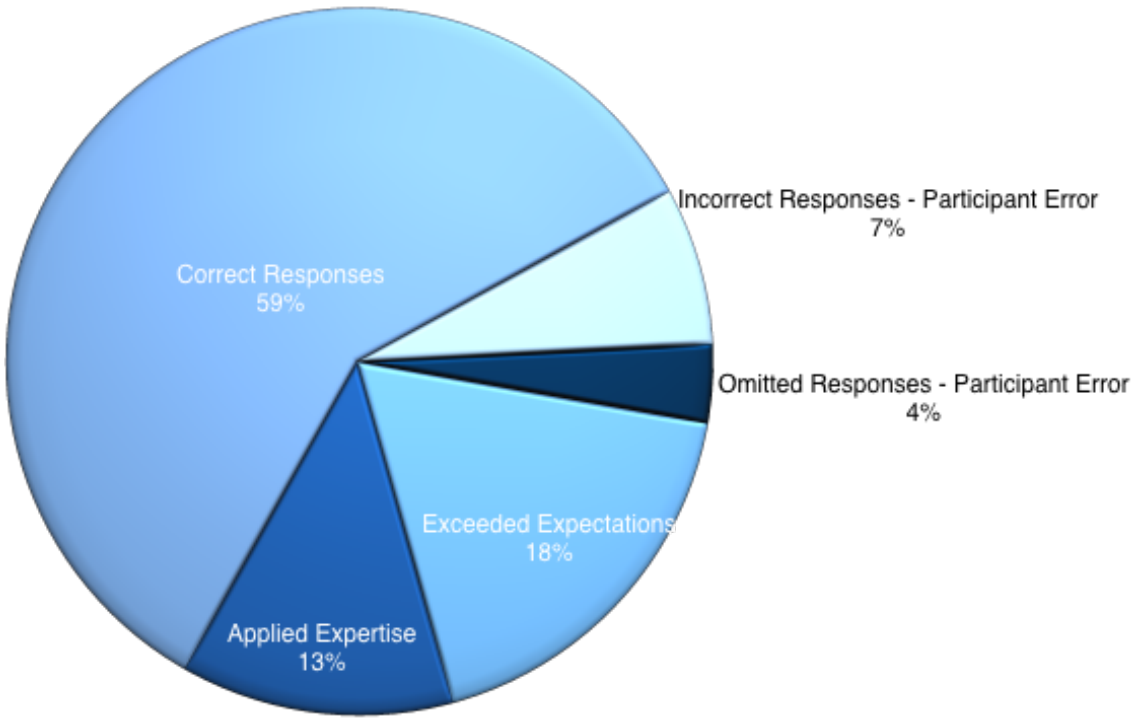


Chart 26: Overall breakdown of results for question 2 across all participants

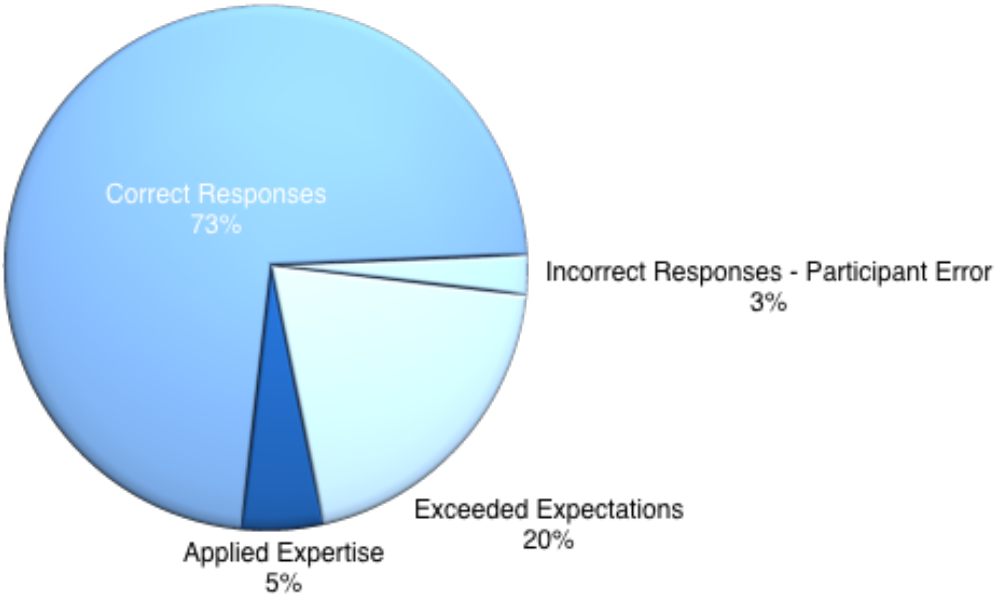


Chart 27: Overall breakdown of results for question 3 across all participants

Chart 27 shows the overall breakdown of results across all participants for questions three. In this case, there was a higher proportion of correct responses at seventy-three per cent, with a consistent quarter where participants exceeded expectations or applied their own expertise. There was a comparatively small number of incorrect answers, at only three per cent. This higher level of correct answers may have been due to the use of a multiple-choice exercise.

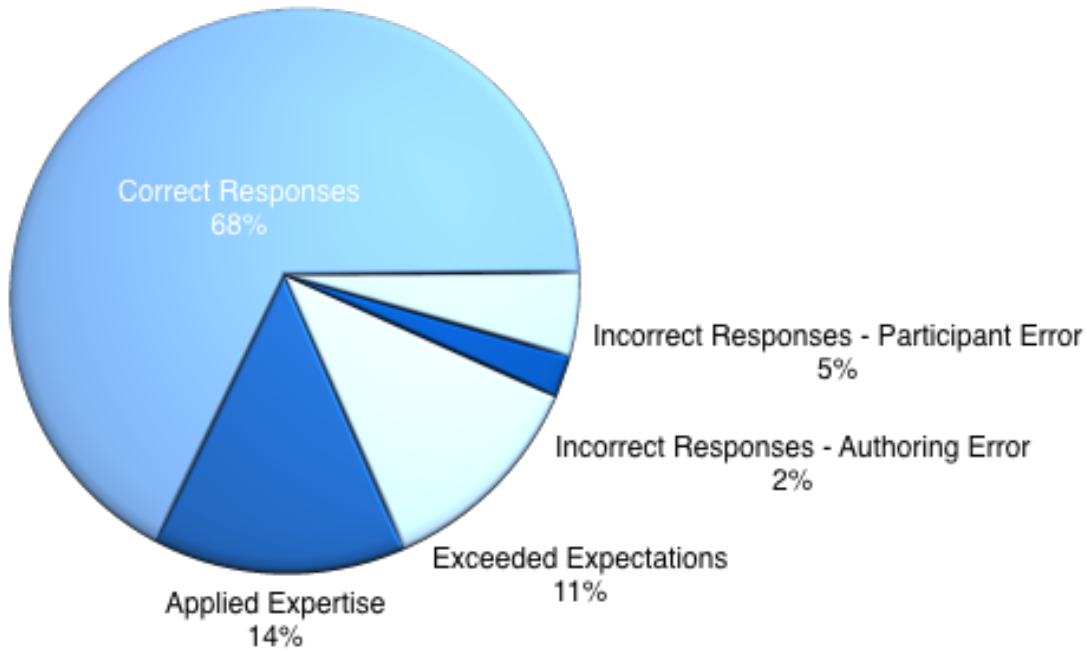


Chart 28: Overall breakdown of results for question 4 across all participants

Chart 28 shows the overall results across all participants for question four. Again, a quarter of the responses were examples of participants exceeding expectations or applying their own expertise. Nearly seventy per cent of the responses were correct and only seven per cent were errors.

These results are further considered by comparing the number of omissions with the numbers for participants applying their own expertise and exceeding expectations.

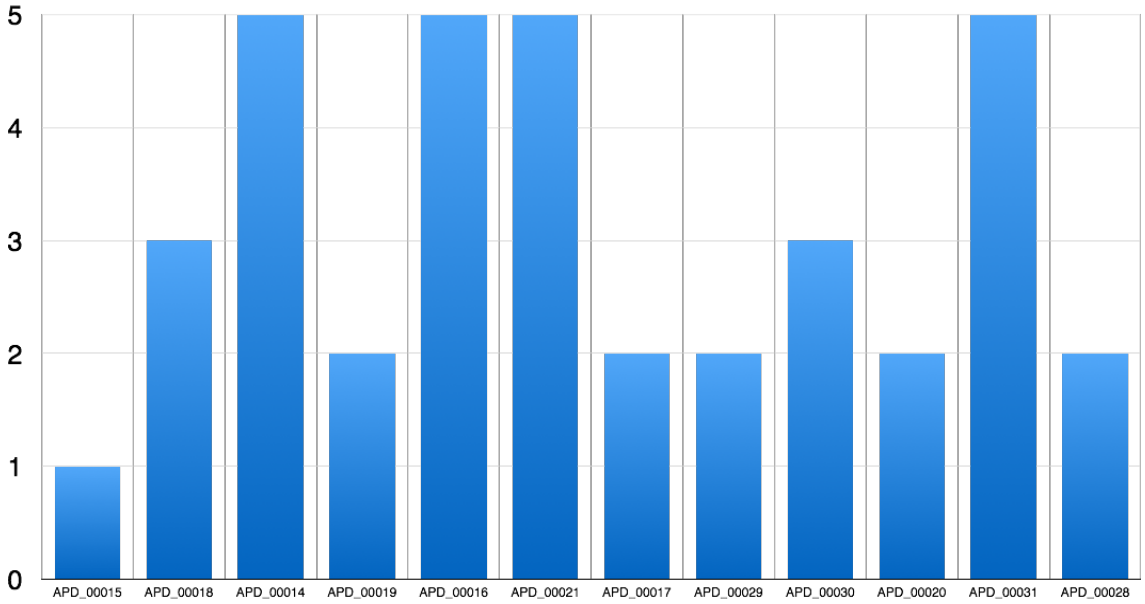
General Points for Experiment 2

As with experiment one, the number of correct responses was significantly higher than the number of errors or omissions, suggesting that *keibi* was an effective tool in guiding participants on how to behave with information based on their answering questions. There was also a notable number of instances where the participants exceeded expectations or applied their own expertise: this showed that they were able to "think outside of the box" and were, for the most part, not misled by the tool, particularly when the original policy excerpts were not necessarily ideal or clear.

Given the number of omissions and errors in authoring policy items, it should be noted that these rarely caused participants to answer questions incorrectly. One participant felt that the information collected was too detailed (or "granular"), and the fact that many of the omitted Elements were not critical to answering the questions suggests that the tool allowed for levels of detail that were beyond the scope of the experiments. It was also clear that there were two classes of information needed: generic rules that would govern all instances of, for example, an information asset like a USB key, as well as rules that should be applied to specific instances, like a particular USB key. Currently the tool allows for both cases in the same screen for a Safeguard, and whilst there were no errors caused by this, a couple of participants mentioned that it might be useful to separate out the generic rules from the specific.

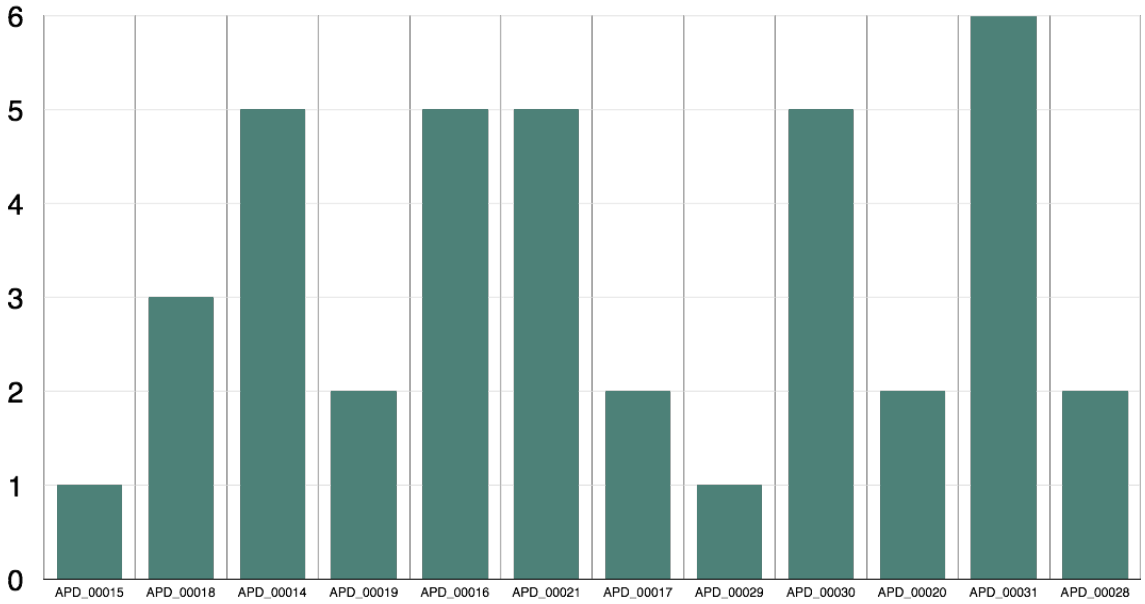
Appendix 20. Questionnaire Scores and Transcription of Group Feedback Sessions

Question 1: Overall, I am satisfied with how easy it is to use this system.



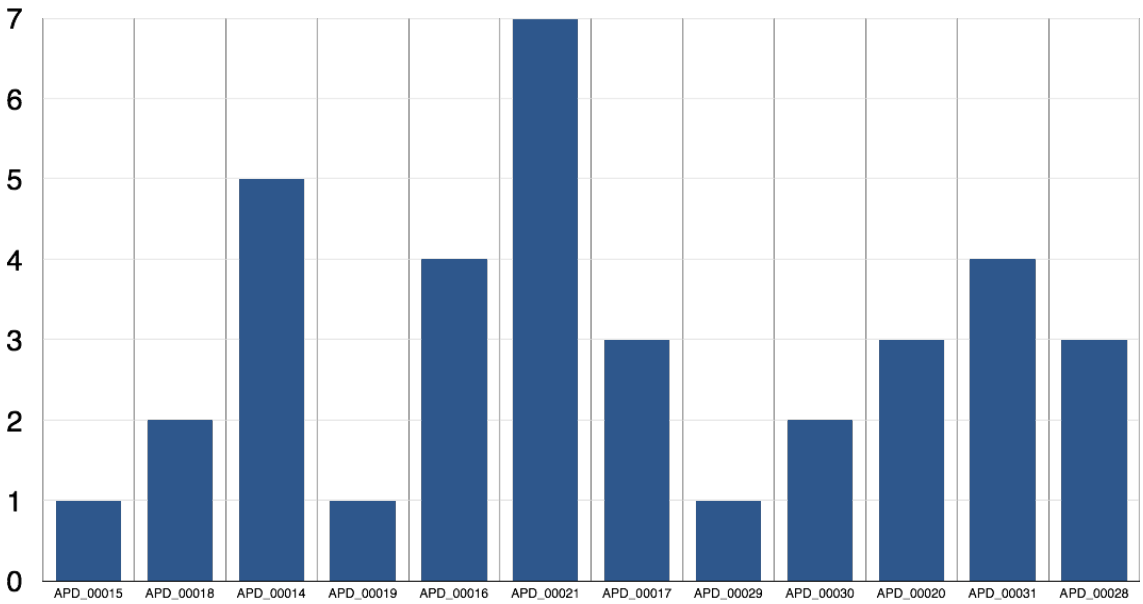
Graph 86: Agreement scores for questionnaire question 1

Question 2: It was simple to use this system.



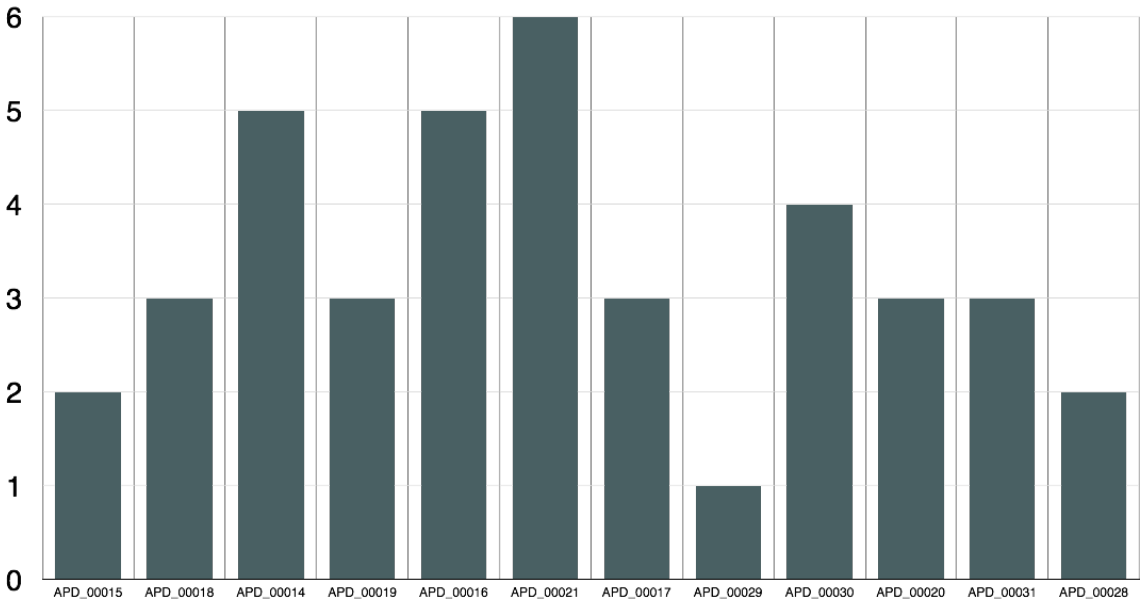
Graph 87: Agreement scores for questionnaire question 2

Question 3 I could effectively complete the tasks and scenarios using this system.



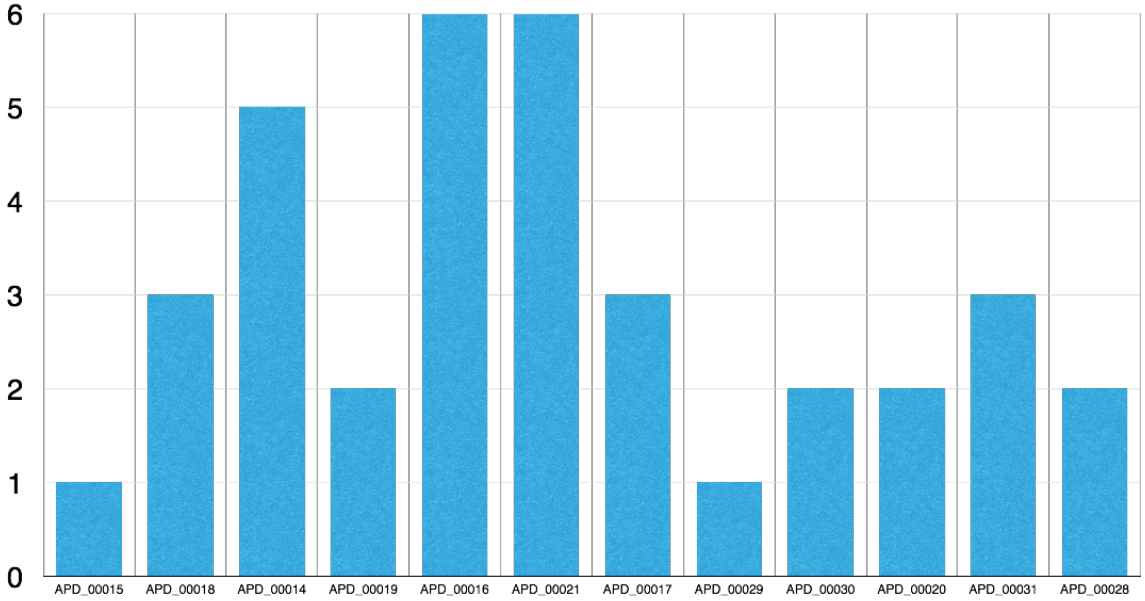
Graph 88: Agreement scores for questionnaire question 3

Question 4: I was able to complete the tasks and scenarios quickly using this system



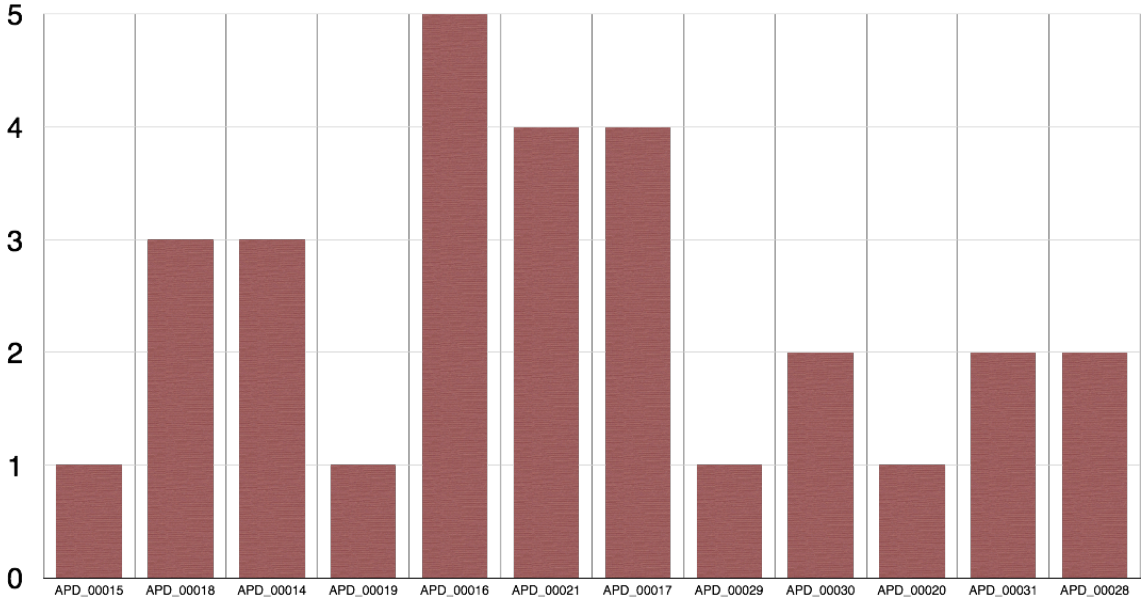
Graph 89: Agreement scores for questionnaire question 4

Question 5: I was able to efficiently complete the tasks and scenarios using this system



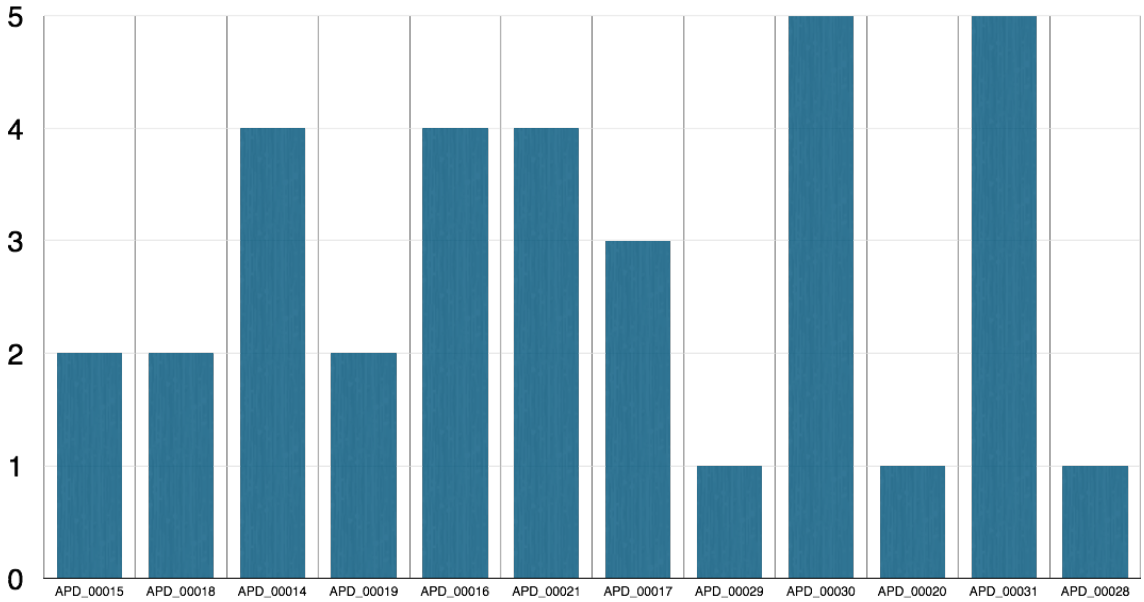
Graph 90: Agreement scores for questionnaire question 5

Question 6: I felt comfortable using this system



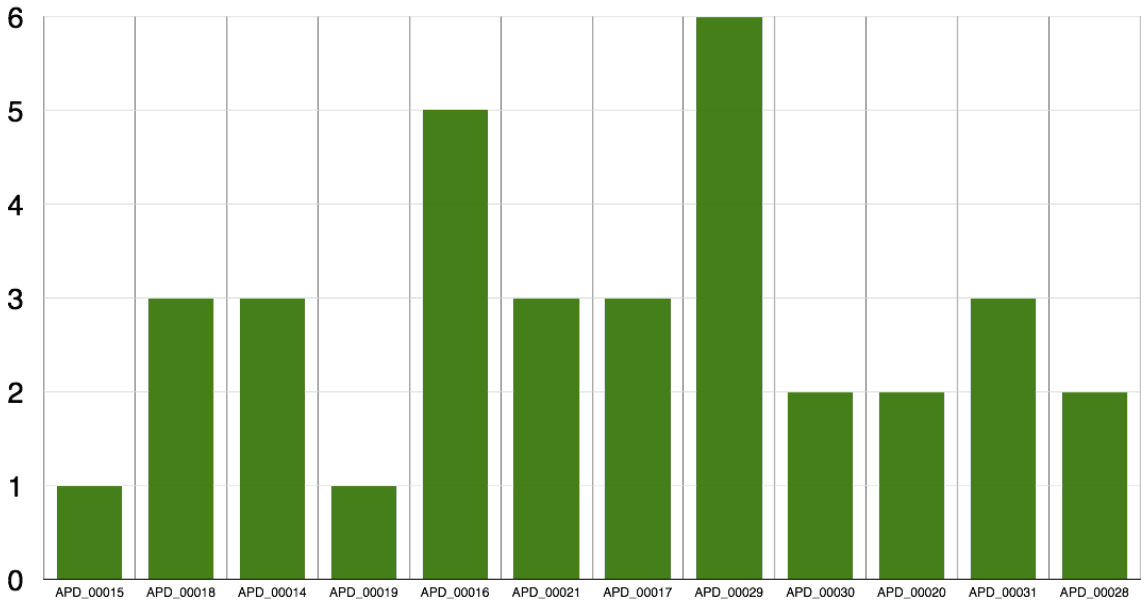
Graph 91: Agreement scores for questionnaire question 6

Question 7: It was easy to learn to use this system



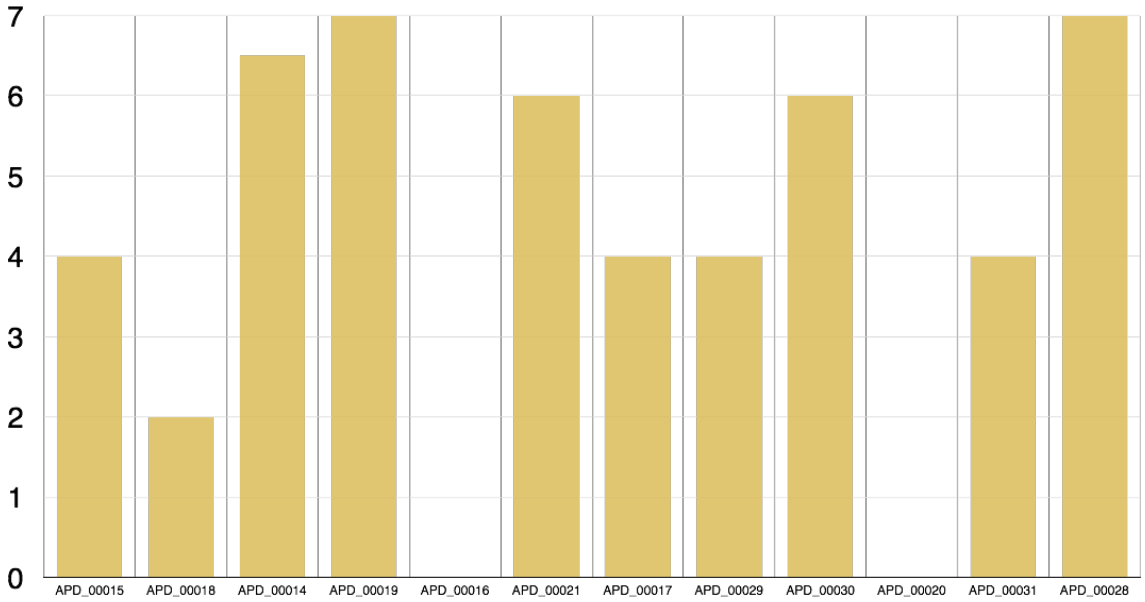
Graph 92: Agreement scores for questionnaire question 7

Question 8: I believe I could become productive quickly using this system.



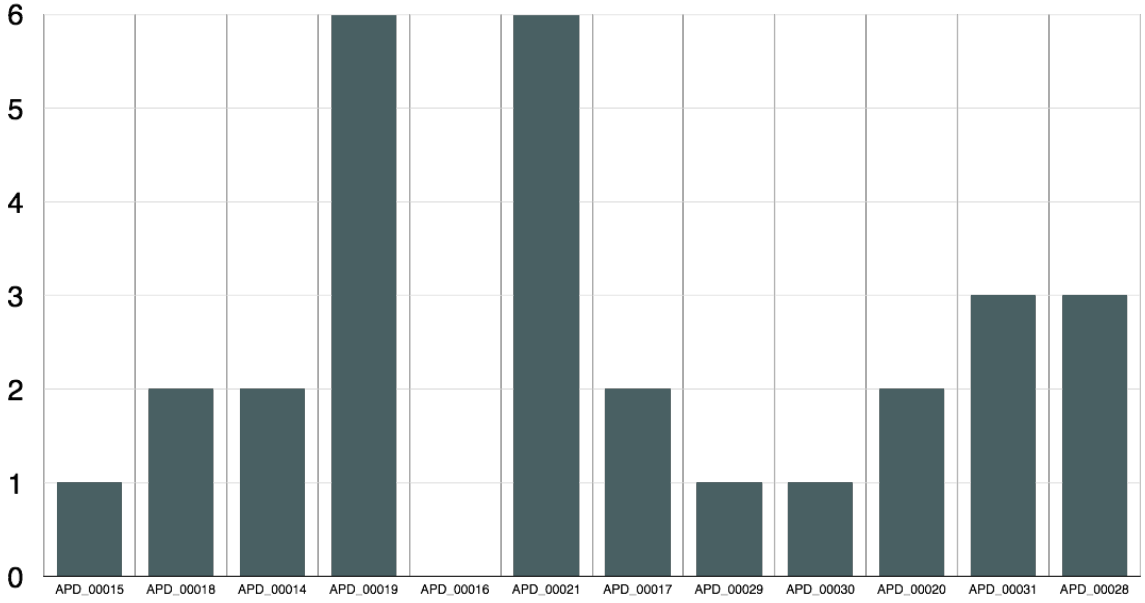
Graph 93: Agreement scores for questionnaire question 8

Question 9: The system gave error messages that clearly told me how to fix problems



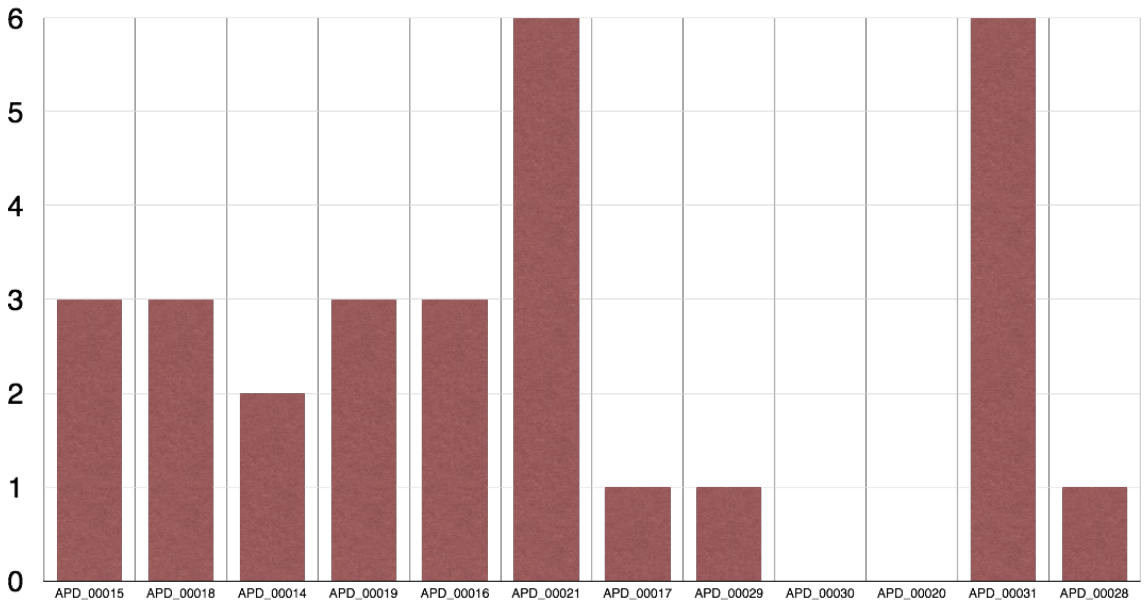
Graph 94: Agreement scores for questionnaire question 9

Question 10: Whenever I made a mistake using the system, I could recover easily and quickly



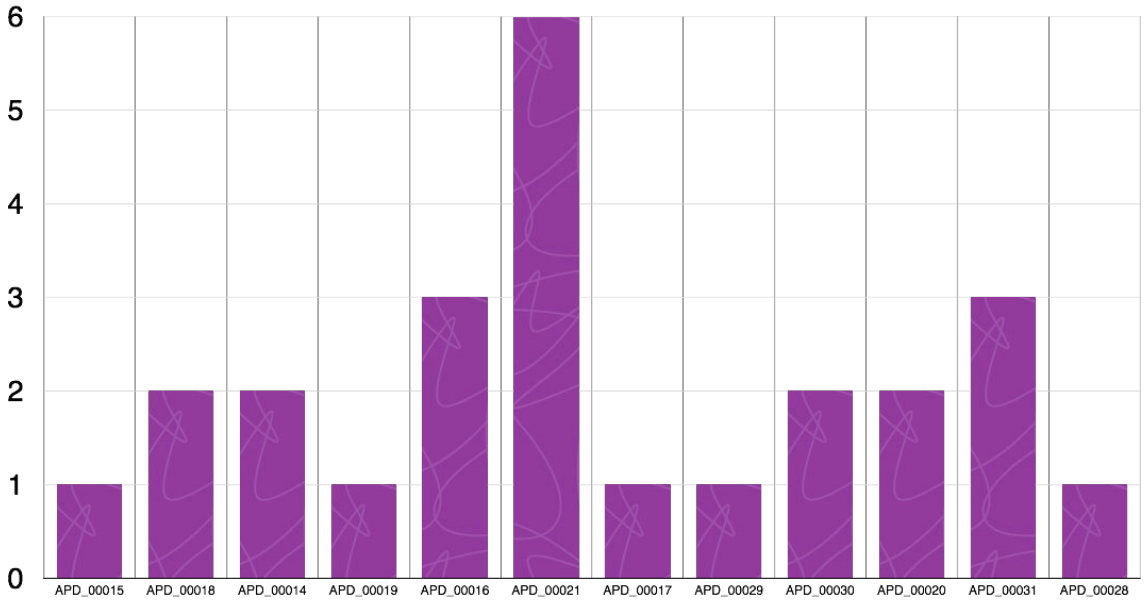
Graph 95: Agreement scores for questionnaire question 10

Question 11: The information (such as on-line help, on-screen messages and other documentation) provided with this system was clear



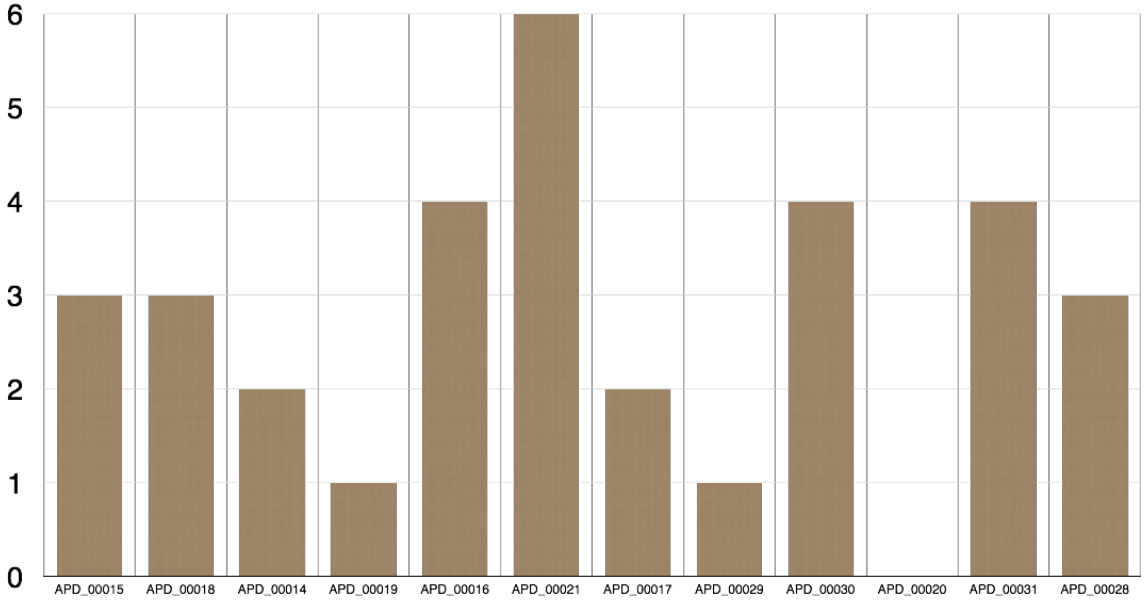
Graph 96: Agreement scores for questionnaire question 11

Question 12: It was easy to find the information I needed



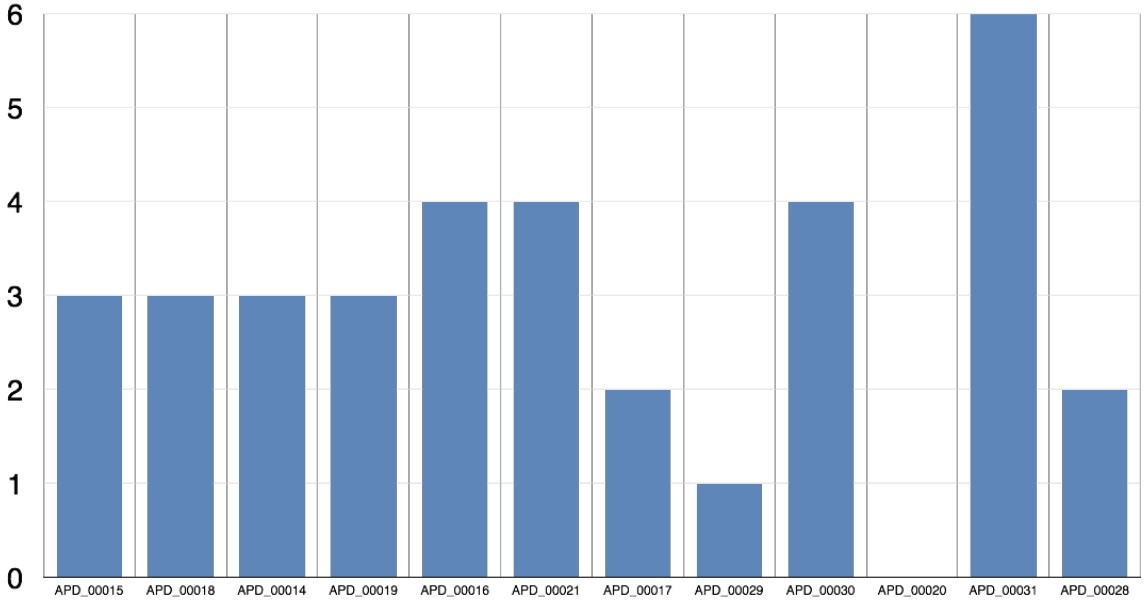
Graph 97: Agreement scores for questionnaire question 12

Question 13: The information provided for the system was easy to understand



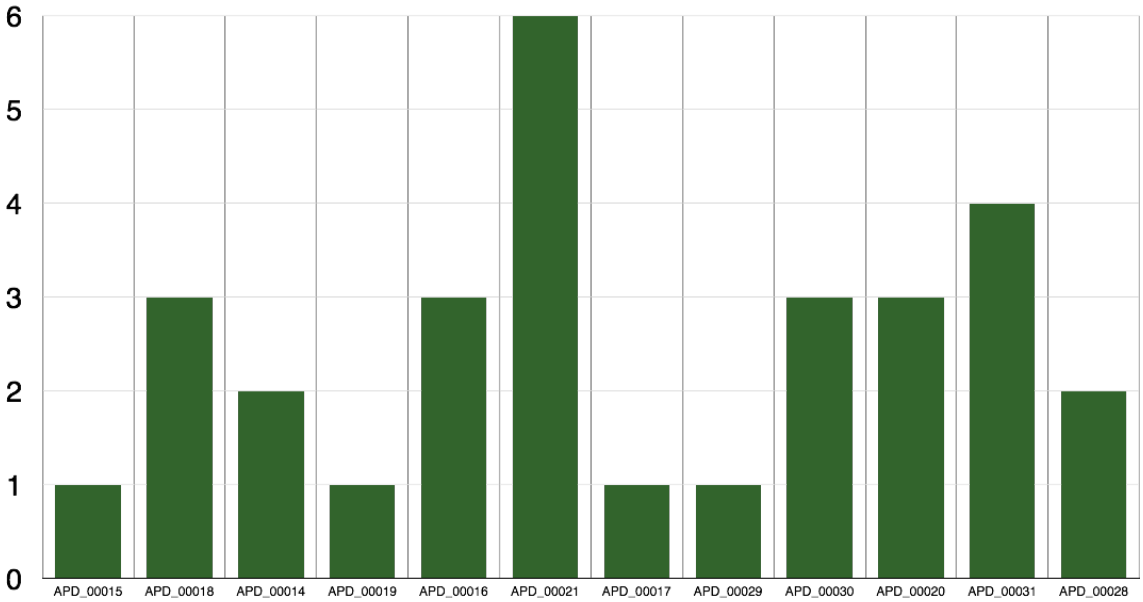
Graph 98: Agreement scores for questionnaire question 13

Question 14: The information was effective in helping me complete the tasks and scenarios



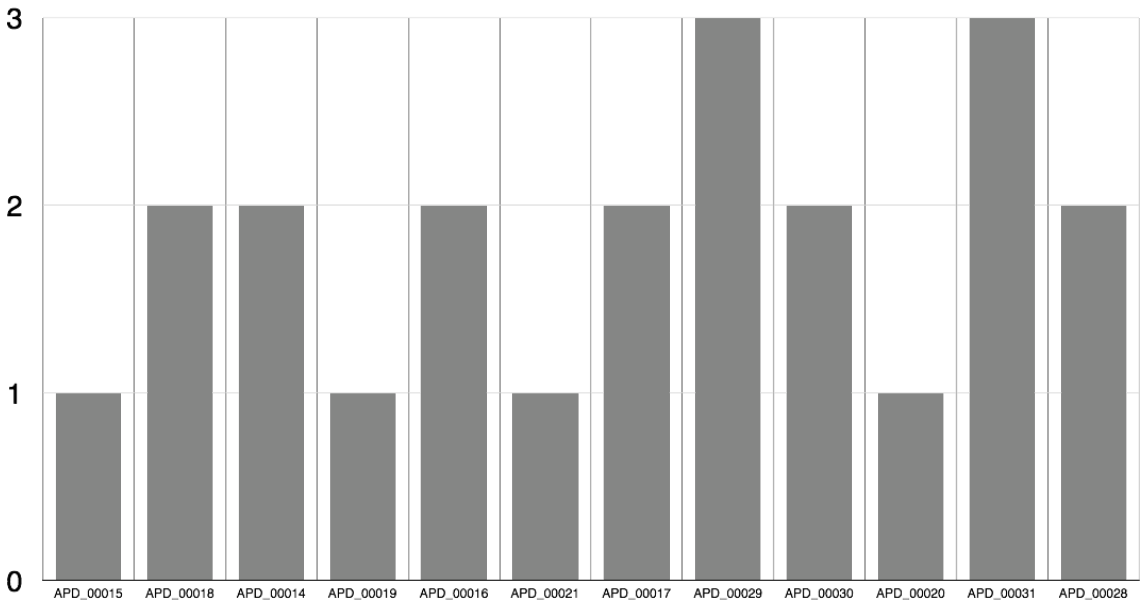
Graph 99: Agreement scores for questionnaire question 14

Question 15: The organization of information on the system screens was clear



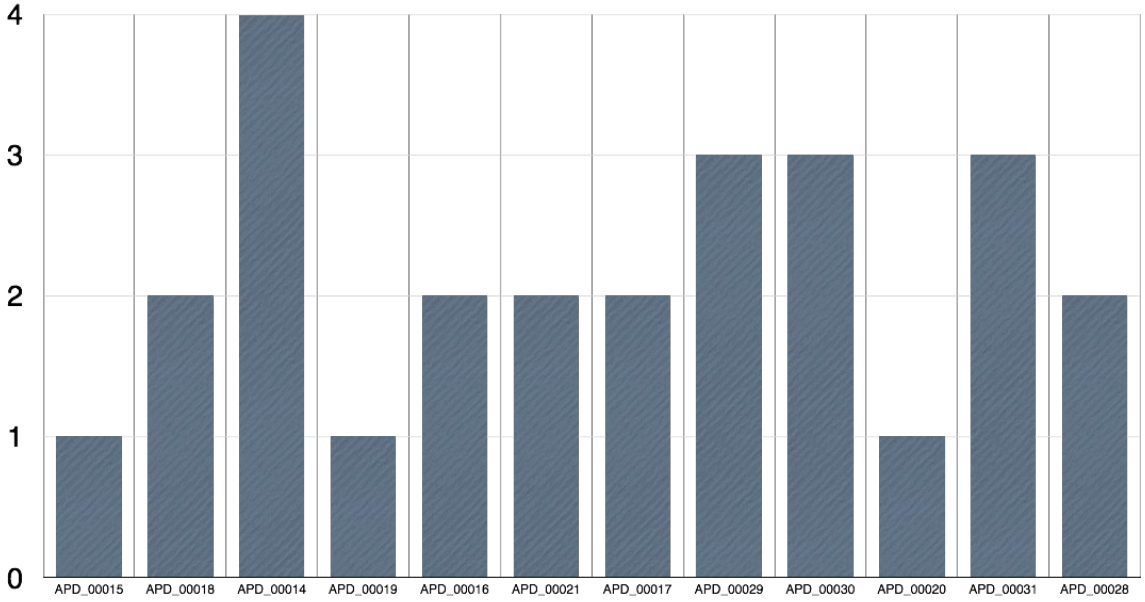
Graph 100: Agreement scores for questionnaire question 15

Question 16: The interface of this system was pleasant



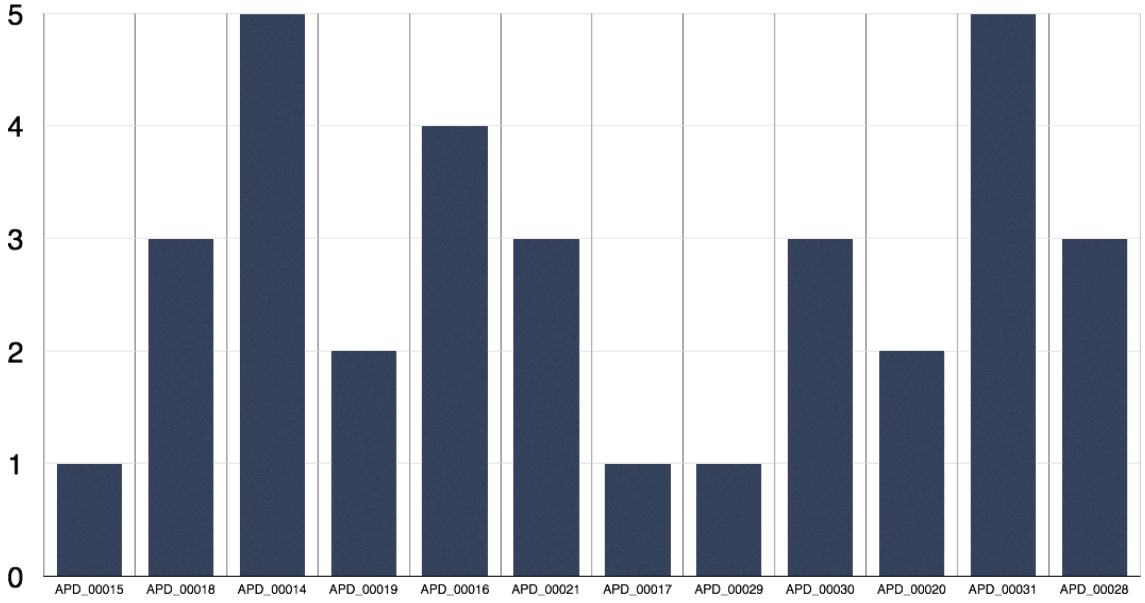
Graph 101: Agreement scores for questionnaire question 16

Question 17: I liked using the interface of this system



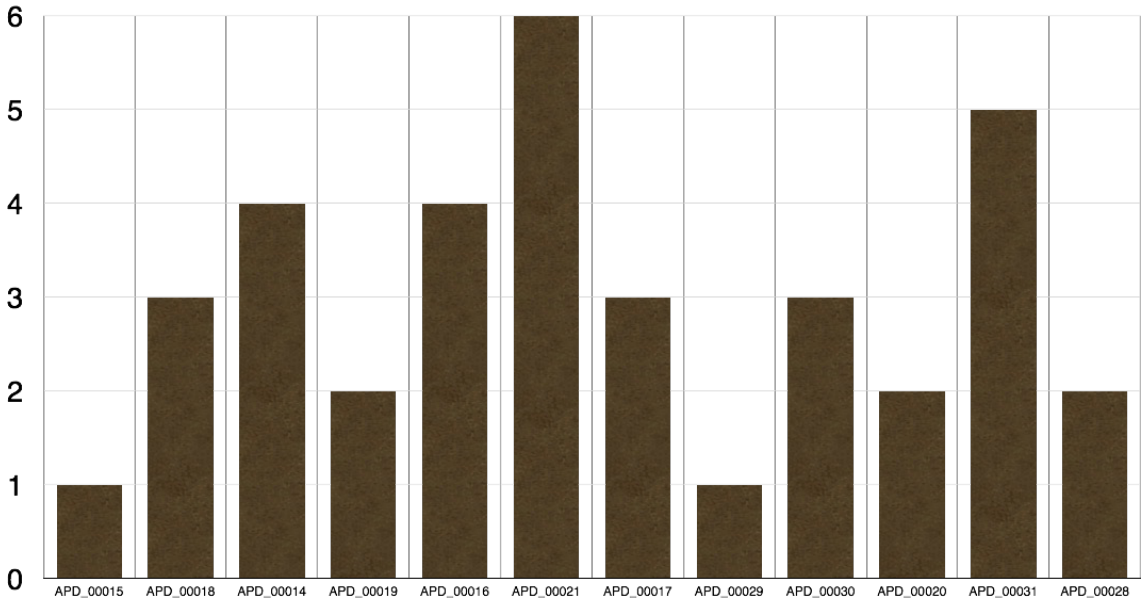
Graph 102: Agreement scores for questionnaire question 17

Question 18: This system has all the functions and capabilities I expect it to have



Graph 103: Agreement scores for questionnaire question 18

Question 19: Overall, I am satisfied with this system



Graph 104: Agreement scores for questionnaire question 19

Group Discussion Transcriptions:

Response to Group Discussion Question 1: What are participant views on the system?

00018 felt that it was hard work to use the system initially due to their trying to understand the new concepts that the evaluation had introduced. They did not quite understand what the concepts meant, for example the way the assets were used, and there was some difficulty distinguishing between activities and safeguards. However once they got used to the system and had used it, it was much easier to understand how the concepts were being represented and how they fit together.

00015 agreed, stating that like using any new tool you had to get your head around it. It really started to seem intuitive when they started to attempt the second example, where there were examples already in place and some information assets had been applied. 00015 said that anyone would want this at the start of a research project. They felt that they did not have a chance to think about how to construct policy in a work environment but expected that this would

work very well. 00015 and 00018 felt that the interface was quite pleasant and liked the colours and layout and agreed tool was quite pleasant to use.

00014 thought there was a steep learning curve to begin with and had to think about how the different concepts / screens / Composition each related to each other. The biggest thing was you had to switch contexts, and even had to drop out of what you were doing to access the help screens. It was easy to click out of something and lose everything you had been working on. Also, they found that the detail for each of the concepts was too granular, and that users would probably end up banging in lots of levels of granularity unnecessarily in the absence a hierarchy of granularity. 00019 agreed with these points. 00019 also felt that they had done well the fact that the investigator had deliberately told them little. 00019 was uncertain as to whether there was a one to one relationship between the excerpt and the items that would be entered into the system (for example, would one excerpt be equivalent to one Safeguard?), but once they knew that this was not the case they knew what to do and it was fine. Both participants would rather have had a consistent granularity across everything. There were also come UI issues: 00014 felt that the UI should allow people the ability to add Information Assets whist they were adding a Safeguard, for example. 00019 suggested that a new asset type or metadata format within each of the screens.

Participant 00017, having filled in lots of online forms, found that this was a bit trial and error, however in comparison to IRAS, which has unhelpful guidance, the help here is very good. The other participants 00016 and 00021 agreed that it was trial and error and see where the details that they entered best fitted. One participant did not know that they were allowed to access help screens. They also felt that coming out of the editing screens to view the help screens was a nuisance - they wanted to crack on with the exercise. They felt that they were also bringing a set of preset views gained from previous advice, which can be misleading. One participant felt that the help screens were a bit wordy, but good.

The participants felt that it was difficult to see where the concepts are from: they were not sure whether something was a safeguard or another one of the concepts. They agreed that their opinion would drive what they entered, unless there was something that would guide them on how to update materials. They

found the Use Context summary useful, where you get a nice summary of the different concepts, but a user could just scroll through those rather than knowing where to look and enter details. One participant got caught out: having entered everything correctly, they had not saved the details and lost them because they had to go and look at a help screen.

Participants 00028 and 00031 felt very unclear at first as to how to encode governance policy into the format provided by the tool. As they went through the questions, they began to understand how the various parts went together. They felt there was something in the tool that seemed useful but not sure they knew what they were supposed to be doing. 00020 appreciated the brief introduction and didn't look at the help sections. 00031 felt that the help section repeated terms that they didn't understand and wanted a worked example. 00028 felt that there was no conceptual framework as to how it all worked together. User interface didn't feel like it was leading a user through in the expected order. 00031 also felt that having selected the Use Context, to view the different elements you had to jump to the left of the screen, and this felt unintuitive.

Participants agreed that, on reading the help screens, tool tip help might be useful. Would you expect people to use it in this way? 00028 felt that there would be the potential for different conventions between a large number of users and that you would end up with a mess if they are all following their own conventions for adding things like labels according to their own choice if they are not consistently guided. Across a more corporate environment this could cause issues, particularly in larger departments. This participant started building own labeling conventions. 00031 did not feel that all the fields were necessary. They found this a bit confusing, and were also not clear on how the concepts related to each other. They also felt that the tool needed rebranding and a better logo.

Participants 00029 and 00030 really liked and enjoyed the idea of having such a tool for interdepartmental use. They felt it was clearly in a in a piloting version and this is an ongoing process. The tool as a tool did exactly what it should do, but there are a few aspects that could evolve in the future. For example, they did not enjoy the user interface that much and felt that a settings section for personalisation would be good, for example to set a different colour background.

Although the help section was straightforward and pretty clear, it could go one better by adding visual material and reduce the wordings. One participant recommended approaching a designer to help with this.

Both participants felt that navigation was fine - the tool is designed to do a very specific task and fulfils its purpose. They felt that the navigation was pretty simple: the main buttons say what each one does. In terms of getting it to a real commercial product, it needs to have something to kick it off. A designer would help. The tool does involve a lot of typing, and the participants felt that it would be helpful to reduce the reliance on this and have more selection boxes and tick boxes. They felt a tiny question mark sign would be a good way to provide information on how to use the tool and what to put in the different fields. The participants felt that a quick start guide combining a visual representation to give the general idea about the core concepts of the system would be helpful. Participant 00030 felt that the help screen was clear, but that it would be easier if you have a button with a popup screen when you select it.

00030 felt the tool was pretty easy to learn, but you need to learn some elements that are not intuitive, for instance they were not clear on what details to put in the different concept screens. Some worked examples would be useful, and 00029 felt some tips on how to write a policy would be useful. They both felt that the workload was mostly authoring. When using it for finding details on how to behave with information, they felt it was intuitive.

The participants wondered how could this work in a department. One or two people may put in policy details, then make these accessible to the users that the policies would govern: The authors should really know how to use the tool properly so as to avoid confusion. The participants agreed that there were two kinds of users. The training requirement would be different for each class. The key role is whoever enters the information and 00030 found this easy to do.

Responses to Group Discussion Question 2: Would they use it in their working practice?

00015 claimed that they would, and 00018 felt that their team already had a lot of policy documentation where lots of the details in keibi were already recorded.

They were not sure how all this information would be put into keibi, and how in the real world they would distinguish between different uses and how to structure it because there are pages and pages of existing documentation. If it were a summary of a policy that would be very useful, but if it is a straight copy then it would be huge task and a duplication of effort. They would not be able to structure the information because they have had to develop a lot of policies already, but keibi would definitely be useful if they did not have to structure any of the existing documentation and having keibi provide a summary of the existing documentation it would be very useful.

00014 replied that: at the moment, partly because service isn't that mature yet, they would not use the system in their working practice. They are just at the point of having policies and having users sign up to them. At the moment, the service does not fit into model proposed by keibi. Both participants felt that, particularly with the queries at the end, they came up with responses that showed what they thought they would do rather than anything they found in the policies authored in keibi. They felt that it would have been nice to have been able to tie together relevant details for given situations and not to have to go and look around what was available to find it. For instance, some kind of metadata that would have tagged particular Safeguards as being relevant to data transfers, using that approach to tie different concepts together in that way that way. Also - they felt that what would be nice would be a kind of question and answer service for researchers to query keibi about what they were trying to do so that it could guide them through the policy details.

00019 raised the point that it would be good to have 'generate policy document' functions so that relevant details could be assembled in, as 00014 suggested, a PDF document or something that could summarise this. They recognised that information governance protection mechanisms are moving away from an era of vast policy documents with numerous pages. What is really needed is something that is tailored to individual users and their requirements, and a mechanism to extract the data for you - key facts, highlighted points and so forth: a mechanism to tell people what they need to do and help with better practices. For example, there are very few cases where researchers actually need identifiable

data even though they feel more comfortable having access this. Venn diagram of requirements and needs for researchers.

00016 felt that it would capture elements of governance and security for external review, new staff and coordinators and setting up a new research study. Another participant felt that If it functions as a support and advisory system, it could be actively used, looking at what is set up and areas of non compliance. It would be helpful to have a wizard, which would make a nice connection between the various concepts and help to cover what is needed for effective governance: there is so much going on and changing rapidly in the area that to have a checking mechanism to make sure requirements were covered effectively would be a fantastic facility.

00016, 00017 and 00021 could see using a system like this if it fitted in to the working practice of an organisation and had all staff "singing from the same hymn sheet" when they accessed different systems. Some form of onscreen guidance would be useful- perhaps more interactive than the static help screens, such as help boxes next to the different text areas. This prompted the author to ask the question: would you like to be trained in using the system? 00016, 00017 and 00021 agreed that they would prefer not to be: there are worse systems out there in terms of usability and intuitiveness - on screen guidance should be enough, provided it was concise and easy to read.

00029 is a data analyst and felt that they had no idea about policies and how find out about them other than asking and word of mouth. They felt that such a tool would not only guide them through and secure their way of working, but would also help them not to make mistakes. They discussed what are currently provided as guidelines - the tool would really help, limiting the need for users to go through pages of documentation, which they felt nobody does. Just by being invited to participate, both participants managed to find documents about security and governance for their project getting ready for the evaluation session, but this is the first time they had done this. Both participants felt that it would helpful to have an indexing and search function so that the right Safeguard is found for a given situation - there is a risk that the wrong Safeguard would be consulted and

offer incorrect advice. 00030 said that if someone has filled in all the important information, would use as guidance.

Responses to Group Discussion Question 3: Can they think of a time in the last year when they might have used it?

00015 does not usually work on projects that involve security policies, but they are now in the process of developing a system that is capturing patient data and they feel that keibi would be useful._00018 felt that a summary for different groups of users would be really useful. They were not actually sure how many people read the existing policy documents, despite being sent the materials. 00018 felt that the *keibi*'s auditing features were important, allowing proof that users had actually viewed policy details. This would be particularly useful for NHS IG toolkit compliance, where currently there is no record that team members receiving policies had read their emails or looked at attached policy documents; having a log where it has been read would be really useful.

00016, 00017 and 00021 agreed that they would have, in one case very recently. When computers went down users had no access to their clinical systems for two weeks. This prompted a change in working practice, including having patients coming in with details written down on old forms. This meant patients were taking a large amount of identifiable data home. This prompted a lot of confusion about handling that situation in terms of information governance, and having access to a readily available set of guidelines in such a case would have been very helpful. The group acknowledged that the system itself would have been inaccessible in this case, but felt that having a printout of policies and procedures with detailed guidance would have rectified this issue.

00014 – could not answer the question directly at the moment because the service is too immature. 00014 can see a time in the future with further development on the tool when it could be useful. Contextual help would be very useful. A feature where you could summarise pertinent safeguards would be very helpful - like a policy generation tool where important information could be put together. Perhaps a policy editor mode and consumer mode, which would summarise and provide feedback on what is needed, tailored to individuals.

00019 felt it would again be useful in the future - the model they operate is pretty simple, working with IDHS and Farr, there might be use. Most of the policy items are burned in the project managers' memories. Having other users knowing about policy items rather than just the managers would be helpful. Even though users are currently sent policies and told to read it, they start behaving as they do with the data, but t doesn't mean they follow the policies.

00020 would not and had not felt the need in the last year mainly because they we're not clear on how the concepts related to each other, so not in it's current form. They though it could be useful to create a knowledge base and decision support for users so they don't have to wade through policy documents, but they felt that this the tool was at the moment purely for encoding. This is the benefit. 00028 felt that it would be a huge job to put all existing documentation in. They did feel that it would be great to have something like this for Principal Investigators. It would also be useful to gain Information Governance Toolkit compliance. Participant 00031 said that they preferred to steer clear of anything that requires complicated data governance. They tend to run projects involving research students that would have to implement policy, but this is not clear when the requirements are more complicated. They did not feel that they fully understood the functionality of the tool and how one could use it. When someone is authoring a policy, there seems to be a presumption that they are starting from scratch. They also wondered whether keibi could produce a printout of policy stipulations? This could be useful for people to use. The participants did express concerns about the encoding, particularly if people relied on keibi where this presumes the encoding has been done correctly. They recognised that it would be time consuming to enter the detail.

Participant 00020 already has a large policy set, but what keibi could be used for, which is required, are Standard Operating Procedures, which could be specifically relevant to the original documentation that is specific to the users themselves. These are quite bulky documents and take a bit of wading through, often are not the best things to consult, but Safeguards could just be SOPs - a way of encoding specific instances and reaction to those instances that are relevant.

Documentation for a research project inherently produces a checklist effect that Principal investigators tended to ignore it. PIs could however produce a data management plan using the details captured by the tool. It would also help to write policy. A wizard would be helpful to develop a Data Management Plan and policy - an important step because DMPs are becoming ubiquitous. With this basic structure, help with IGT and ISO 27001 compliance would be possible, and this would help to lead users to the same conclusion. This could also serve as an effective document management for PIs. Collaborative would also be useful: the details would be in one place, and the audit trail is big advantage. This would certainly assist with Staff induction. Training and education, you need a bit of framework, template or wizard on top. Have people take something away.

If there was a way of commenting on improvement plans, if compliance could be done on spread sheets, encode 14 information governance controls for secondary use, you still have to record what you do. APD00031 felt that you should only see what you need. APD_00020 felt there was a hierarchy to those Safeguards, but this may be a fluke as to how they've been entered. Don't share - conflicts and arrangements not to share but later you can if you do this. Some that encapsulate. May want a glossary. Control is used - translation to different nomenclatures. Controls in context.

00029, 00030 and 00031 did not feel that there was a particular moment they were in doubt of what they needed to do, a general knowledge of the tool would be helpful. They both agreed that when interacting with another organisation, or helping students, they would like to advise the collaborators on how they should behave, on what to do and what they should know. 00030 felt it would be useful for teaching purposes for teaching purposes.

Responses to Group Discussion Question 4: Did the tool make participants think about issues surrounding information security and governance?

00015 and 00018 could recognise everything in the tool and did not see that there was anything missing. They both found were reassured to see that the information

governance and security requirements of policies seemed to be present in the tool. They also felt that the tool would be a good educational resource. 00015 said that the only time they came across policies was in projects that involved the data sharing of social care records. These in isolation are pointless, and something more rounded for the context of data sharing would be needed. keibi would be very useful to orientate a new member of staff on a research project, for example, as a means for them to become familiar with the requirements for safe and secure working practice within a project. Both agreed that keibi would be really good as an educational tool and should fulfil this purpose.

00014 felt that the tool encapsulated everything that they knew about the area, breaking it into categories. They felt that one has an holistic view in ones head, and they didn't tend to think about it in terms of the concepts as they were presented. They pointed out that humans learn holistically, not in a tabular form. In general, 00014 does not tend to think about the information security and governance concepts as separate, and it felt a bit artificial breaking it out like that, which is what tends to happen when you place various concepts in a database. 00019 agreed with these views. They also felt that with the last set of questions, there was a huge mismatch between what the policy says and what you actually do with what you can practically work with: there is always a mismatch between how much users bend the rules when compared with what can be realistically applied.

In answering this question, both participants 00014 and 00019 considered the challenges that face organisations that process sensitive information. There is a "grey area" of whether information assets are identifiable and fall outside of pseudonymised classification, making them covered by the Data Protection Act. This can make determining how best to handle this information difficult to know. An example of where things could go wrong was a recent discovery of a USB key that had patient photos that could not be identified and were considered anonymous was found on a nearby road from a hospital, but there could still be an impact, particularly if you apply the "Daily Mail" test. This would lead to an organisational impact by association, undermining their reputation for handling sensitive health information. The participants feared that, whilst there is a constant fear in people's heads about security breaches, until it happens there is

not quite the same motivation to manage these issues. The participants also recognised that there are varying degrees of sensitivity between medical conditions where particular concern about issues is warranted, where a hip replacement may be less sensitive than a record of a venereal disease. The Participants identified insurers as being a group that cause particular concern for people, where lost or misused records could have an inverse impact on their cover or its cost, though they accepted that a company may not be able to legitimately use information that had come to them through an unauthorised route. An additional issue is that how best to proceed is never clear cut - there is always more than one way of protecting an asset. A useful example from the guidelines is the example of forbidding the storage of even anonymised data on a CD ROM, but guidelines don't think about the unknowns in general. The context of use can get particularly sticky, where on the one hand you can get an expert in to think about specific factors and proposed solutions, but the result is always ten pages of narrative guidance of what you shouldn't do. But what should data users do - for instance send information by email encrypted in breach of policy and not tell anyone? This breeds bad practice. Guidelines on how to write effective policy would be particularly useful.

The participants considered how keibi might be able to help handle these issues. 00014 felt that if keibi were fully populated and handled small numbers of policy items, it could perhaps serve as a decision support tool and provide heuristics to detect potential exposures, applying a red, amber green indicator of exposure based on what a user were trying to do, this would be a positive step in the right direction. For example, if it's red, that is a clear indication that a user should not attempt to do what they are planning. Participants agreed that keibi could also serve as a means to handle the SLMS risk assessment tool in a more usable format other than an Excel spreadsheet. This was pertinent because the participants both liked the interface and felt the tool had a lot of potential. They agreed it was usable, but needed a bit more development: whilst it had the foundations in place, there a couple of usability elements, like needing to save items as draft and then come back to them to complete them: it would be best not

to have to navigate away from entering details of a Safeguard to complete another Information Asset or seek help because that interrupted the flow of activity.

00016, 00017 and 00021 felt that using keibi helped them to think about the play off between pragmatism and good governance, particularly where there are failures. There is still a lot of uncertainty about how to appropriately handle sensitive information, and they felt that they hoped that any failures would happen in a limited and not harmful way. There is a slight twinge of pragmatism: there are limits to within which people handling sensitive information can operate.

The group also discussed the issue of people who process sensitive information not always being fully aware of their responsibilities. Questionable practice does happen - more junior staff and students send patient identifiable information over Gmail, for example, and staff sometimes have a lack of awareness. Another example of access to information in the healthcare setting is the ability to look at thousands of peoples' records without a legitimate purpose. This is current practice, but it raises questions about what access should be permitted, how it should be best limited and how the plethora of general rules for accessing and using information can be brought together to a unified and consistent set of guidelines and support that is easily and consistently understood.

00016, 00017 and 00021 also felt that getting governance and security right is becoming especially important given the recent incident with Edward Snowden - there seems to be an increasing pressure to talk to people face-to-face, meaning that more information is now being shared outside of the IT systems that have been created to capture such details. The group asked whether this was unintentionally introducing other risks to protecting confidential information, particularly if discussions were being held in more public spaces like the canteen or even on public transport. There is a balance to be struck, however: for instance - would patients really mind if their clinical data was emailed to them, particularly if there was a benefit for quicker access to information and more convenient communication? Patients themselves could also be inadvertently be breaking the rules and making disclosures, over which health service providers and the research community have no control whatsoever.

00016, 00017 and 00021 agreed that keibi could help to provide guidance and reassurance that good practice is being adhered to - it does a lot of the work for you which is nice, and if it can present the details in a succinct form that would be wonderful.

00016, 00017 and 00021 felt that it created a few thoughts, which they had not thought about before. They offered that it would be helpful to create any guidelines, it would be good to let people play around with the tool, invite feedback and consultation as policies are developed. Participants 00020 and 00031 felt that it did: by making them try and code the rules in some format they realised that the rules were poorly specified. It did make the Participants think about what they different rules were actually trying to achieve and represent.

Additional comments:

00015 asked whether the tool could be used to archive or destroy data when the appropriate usage and retention periods. Whilst it would be possible to author this in keibi it would rely on the information management system to implement the functionality to archive and securely delete data at the appropriate times. keibi could handle the rules and legal bases required to determine the point and execution of such functions, but those functions would have to be developed in the first place.

00018 asked whether keibi can be used within their institution's secure data safe haven infrastructure. They felt that it would be a very useful replacement for an Excel Spreadsheet that held details about information asset usage and management and offered a risk assessment function: keibi was a much more pleasant system to use and it would be very useful to replace the spreadsheet with it. The response was that keibi could certainly store all the details held by the spreadsheet, and that a risk assessment function based on the information captured in keibi was scheduled as further work. The participants asked whether the tool could link to the actual policy and received the answer that it could through the Legal Basis Composition.

00020, 00028 and 00031 were not sure how Activities and Safeguards differed. They also felt that the tool did not actually turn them into a form that is

computable. The participants wondered whether the tool could or would do some more useful decision support. They also felt that flow diagrams would be a useful way of presenting the details in the Safeguards and to show the relationships between the concepts. 00028 suggested that users should be able to see Safeguards by Activity - and offer sorting and indexing facilities for the details.

Participant 00020 had a very similar reaction to 00031 - they have a managerial responsibility for creating SLMS policies, so the level of interest would be a list of Safeguards that they engage at a technical level. They do not engage at the level of specific data items: that has to be done by the people that abide by the policies. The tool does make you think about policy documents - SLMS does have in the order of twenty plus documents. The participant recognised the amount of effort that would be needed to encode all those documents and then apply to specific activities. Being forced to think about it makes you think more, but how practicable that would be in a large scale framework is not clear. They recognised that ISO certification does make organisation think to that level of detail.