Faculty of Science & Technology

Department of Computing and Informatics

Jane Henriksen-Bulmer PhD Thesis

Incorporating Contextual Integrity into Privacy Decision Making: A Risk Based Approach

Abstract

This work sought to create a privacy assessment framework that would encompass legal, policy and contextual considerations to provide a practical decision support tool or prototype for determining privacy risks, thereby integrating the privacy decision-making function into organisational decision-making by default.

This was achieved by way of a meta-model from which two separate privacy assessment frameworks were derived, each represented as a stand-alone prototype spreadsheet tool for privacy assessment before being amalgamated into the main contribution of this work, the PACT (PrivACy Throughout) framework, also presented as a prototype spreadsheet.

Thus, this work makes four contributions. First, a meta-model of Contextual Integrity (CI) (Nissenbaum 2010) is presented, where CI has been broken down into its component parts to provide an easy to interpret visual representation of CI. Second, a practical privacy decision support framework for assessing data suitability for publication as open data, the ContextuaL Integrity For Open Data (CLIFOD) questionnaire is presented. Third, the scope of the framework is expanded upon to include other industry sectors or domains. To this end, a data protection impact assessment (DPIA), the DPIA Data Wheel, is exhibited that integrates the provisions brought in by the General Data Protection Regulation (GDPR) with CI and a revised version of CLIFOD. This framework is applied and evaluated in the charity sector to demonstrate the applicability of the concepts derived in CLIFOD to any domain where data is processed or shared. Finally, this work culminates with the main contribution of this work, one overarching framework, PrivACy Throughout (PACT).

PACT is a privacy decision framework for assessing privacy risks throughout the data lifecycle. It has been derived and underpinned by existing theory though the amalgamation of CLIFOD and the DPIA Data Wheel and extended upon to include a privacy lifecycle plan (PLAN) for managing the data throughout its data life cycle. PACT, incorporates context (using CI), with contemporary legislation, in particular, the General Data Protection Regulation (GDPR), to facilitate consistent and repeatable privacy risk assessment from both the perspective of the data subject and the organisation, thereby supporting organisational decision making around privacy risk for both existing and new projects, systems, data and processes.

0.1 Publications

Henriksen-Bulmer, J., Faily, S., Jeary, S. and Katos, V. Putting context into the Data Protection Impact Assessment: the DPIA Data Wheel. In *Computer Journal*, under review, January 2019.

Henriksen-Bulmer, J., Faily, S. and Jeary, S. Privacy Risk Assessment in Context: A Meta-model based on Contextual Integrity. in *Computers & Security Journal*, accepted for publication, currently in preprint, January 2019. (Henriksen-Bulmer et al. 2019b)

Henriksen-Bulmer, J., Faily, S., and Jeary, S. Implementing the General Data Protection Regulation (GDPR) in the Charity Sector: A Case Study. *Privacy and Identity Management - Fairness, accountability and transparency in the age of big data* (2018), Chapter in Springer Series Textbook: 13th IFIP WG 9.2, 9.6/11.7, 11.6 international summer school, Vienna, Austria 20-25 aug 2018, revised selected papers ed. IFIP Advances in Information and Communication Technology. Springer International Publishing, Switzerland, 2019. Currently in preprint, January 2019 (Henriksen-Bulmer et al. 2019a).

Henriksen-Bulmer, J., Faily, S., and Jeary, S. DPIAs for Charities: a Charity Sector Specific DPIA Framework. In IFIP Advances in Information and Communication Technology (2018), *13th IFIP WG 9.2, 9.6/11.7, 11.6 International Summer School, Vienna, Austria 20-25 Aug 2018*. (Henriksen-Bulmer et al. 2018)

Henriksen-Bulmer, J., Faily, S., and Katos, V. Translating Contextual Integrity into Practice using CLIFOD. In *Proceedings of the 2018 Networked Privacy Workshop at CSCW (2018).* (Henriksen-Bulmer et al. 2018)

Henriksen-Bulmer, J., and Faily, S. Applying contextual integrity to open data publishing. In *Proceedings of the 31st British HCI Group Annual Conference on People and Computers: Digital Make Believe (2017)*, British Computer Society. (Henriksen-Bulmer and Faily 2017)

Henriksen-Bulmer, J. A Framework for Public Bodies for Managing the Secure and Appropriate Release of Open Source Data. In *British HCI 2016 Doctoral Consortium (2016), Bournemouth University.* (Henriksen-Bulmer 2016)

Dissertation Declaration

I agree that, should the University wish to retain it for reference purposes, a copy of my dissertation may be held by Bournemouth University normally for a period of 3 academic years. I understand that once the retention period has expired my dissertation will be destroyed.

Confidentiality

I confirm that this dissertation does not contain information of a commercial or confidential nature or include personal information other than that which would normally be in the public domain unless the relevant permissions have been obtained. In particular any information which identifies a particular individual's religious or political beliefs, information relating to their health, ethnicity, criminal history or sex life has been anonymised unless permission has been granted for its publication from the person to whom it relates.

Copyright

The copyright for this dissertation remains with me.

Requests for Information

I agree that this dissertation may be made available as the result of a request for information under the Freedom of Information Act.

Signed:
Name:
Date:
Programme:

Original Work Declaration

This dissertation and the project that it is based on are my own work, except where stated, in accordance with University regulations.

Signed:

Acknowledgements

First of all, I would like to thank my supervisors, Shamal Faily, Sheridan Jeary and Vasilis Katos for their valuable, insightful and continued guidance and support, it has been an honour, privilege and genuine pleasure to work with all of you.

More particularly, first, I would like to thank you Shamal for the many discussions and debates that helped shape this work, for the opportunities for collaborations with partners such as the Charity and HuaHana, and for the help and guidance provided throughout. Vasilis, thank you for finding time to provide a listening ear whenever it was needed, for inviting me to participate in the Ideal-Cities project, and for your steerage and encouragement. Last, but by no means least, thank you Sherry for reigning me in when I needed it, for listening, directing and keeping me on track when I got lost, your unwavering support, friendship and encouragement and for helping make sure the story got told, Goldilocks made it through the woods in the end!

Second, I want to give a heartfelt thanks to all the practitioners, peers and facilitators who worked with me in the case studies, training sessions and workshops. I appreciate you taking time out of your busy schedules to collaborate with me in trialling my frameworks and providing helpful feedback and suggestions for improvements. Thank you for the questions, debates, deliberations, suggestions and advice that helped shape this work to make the final framework practicable and useful. You helped make the journey interesting, fun and enjoyable and all the hard work worthwhile.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Individuals now have access to much more information through personal computers and the internet than ever before. Indeed; ”the web is considered the largest publicly accessible data source in the world” (Kimble and Milolidakis 2015). As this resource grows, so too does the quest for more data. People continue to want and request more data, particularly public sector information. For example, organisations may use information from the electoral roll to confirm customer details or rely on the live traffic data feed to direct their fleet. Similarly, individuals may want to access information about what new properties are being built in their area (planning proposal) so they can comment or object, or wish to find out how the local authority spends its funding (e.g. budgets or forecasts). Therefore, pressure is on government to release more data, instigated, in part, by pressure groups such as the ’free our data’ campaign (Arthur and Cross 2015); (Sunlight Foundation 2010).

What happens to data and how data is handled by organisations, their employees and individuals depends on the decisions made around that data. Individuals make decisions every day, not just on behalf of the business they work for, but also on all manner of aspects of their lives from the minute they wake up and step out of bed. Many of these decisions are subconscious or habitual; people do something because that is how they have always done it or because it is customary for that thing to be done in a certain way. Some decisions are more conscious. For example, on a personal level, an individual may consciously respond to a comment or post by a friend on social media without putting too much thought into the finer detail or indeed their privacy. However, if that individual is instead submitting a job application, a planning application or writing to the local council to complain, they may think a little more about what they write, how the correspondence is formulated and where this is submitted or sent to.

At the organisational level, data is at the core of most businesses and, in many cases, forms the lifeblood of the business (Daley 2016); (Hirsch 2013). Data and data analytics have become big business; ever larger volumes of data are now collected, distributed and shared at a rate never before seen or imagined just a few years ago (Boyd and Crawford 2012). Data analytics affords exciting opportunities for both research and competitive advantage using the power of data (McAfee and Brynjolfsson 2012), but it also brings new threats and concerns, particular around privacy. Advocates of data analytics claim that data protection rules and guidelines will safeguard us. However, even if privacy is preserved initially, once the data is copied, distributed, linked and shared, can we be confident that privacy will remain intact? It is, therefore, imperative that organisations consider, not just whether data protection regulations might be breached if they share or publish the data, but also take into account the context in which the data was collected, stored and shared and how this might affect any individuals to whom the data pertains (the data subjects).

1.1.1 Domain

These questions are what motivates this thesis, the intention is to establish an effective way to support the increasing uses and sharing of data, while ensuring that privacy has been sufficiently considered and effectively protected by the people who handle the data. To this end, it is thought that open data will provide an ideal starting point for determining how this can be achieved. This domain has been selected because open data is data that is freely available for anyone to use and reuse without restriction (G8 2013), it is perhaps one of the most difficult domains within which to effectively support privacy decision making. Therefore, if privacy decision making can be effectively supported and maintained in this domain, the principles or methods used to achieve this should be applicable to any domain.

There are a number of guidelines and frameworks to support risk and decision-making around data, most however, are practice rather than research led, meaning these have been formulated based on practitioners knowledge and experience rather than on formal academic research, e.g. (FERMA 2003, BS ISO 31000 2009, BS ISO/IEC 29100 2011). Further, most of these frameworks have been created to support risk and/or security with privacy incorporated as a small part of the risk or security consideration to ensure legal obligations around privacy are met (BS ISO/IEC 29100 2011, NIST 2010, FERMA 2003), few have been created specifically for privacy. Of the guidelines that have been created to address privacy specifically, these either provide high level guidance only (Information Commissioners Office 2014, Cavoukian 2011), or they have been created to address a specific field such as policy formulation around open data publishing (Altman et al. 2015). This thesis argues that privacy risk is, and should be, sufficiently important to warrant being assessed and categorised in its own right, rather than as a component part of other risk assessment processes.

However, to understand privacy, a variety of different areas need to be explored first to show how decisions and risks can influence and affect privacy and consequently, how these decisions can influence whether or not data is shared or published in open format.

1.2 Research Questions

This work looks at how organisations currently make decisions around privacy. This will inform the creation of a privacy decision-making framework that will facilitate more streamlined privacy decision-making processes. This will begin by looking at privacy decision making around open data publication and data sharing practices with a view that, as the work progresses, the framework will be equally useful and applicable to support privacy decision making for any data sharing, processing or publication practices. The aim is that the final framework will enable organisations to self-attest to privacy compliance and thus, provide users with assurances that their trust in organisations ability to safeguard their privacy has not been misplaced.

The aim of this work is not to merely create another theory, but rather, to create something that will be used by organisations in practice. Therefore, this work will seek to understand the existing view of the world on privacy and ”rather than simply examining nature” this work seeks to help ”create a view of the world” (Hyland 2004) that will make a difference to practitioners everyday lives and bring about an improvement in processes and thinking.

To this end, this work suggests that organisations should adopt a holistic risk based approach to the privacy decision-making process and incorporate privacy into the organisational decision-making process. This will, for example, enable organisations to make informed privacy decisions before data is published as open data. Thus, this research asks;

> *”How can privacy assessment be incorporated into organisational decision-making in a practical manner, encompassing legal and contextual considerations, to provide repeatable, effective decision support for determining privacy risks and facilitating the integration of the privacy decision-making function into organisational decision-making by default?”*

To answer this question, this work will begin with a review of the literature. This will look at how organisations currently make decisions around privacy and what support is currently available to them to support the decision making function, starting with a look at privacy decision making in the open data domain. This domain has been selected as the starting point in view of the perceived higher risk of providing unrestricted access to open data, allowing uses to download, re-use and amend as they see fit (Open Data Institute 2016, Open Data Charter 2015). This review will therefore begin with reviewing privacy, data privacy and the open data landscape before moving on to look at how data privacy may be protected and organisational decision making for privacy.

Decision making around data privacy is concerned with identifying risks and ways that data privacy can be preserved (Cavoukian 2011, Information Commissioners Office 2014, NIST 2010, BS ISO/IEC 29100 2011, Altman et al. 2015). Privacy practitioners who are decision-makers need tools that offer options and choices of actions rather than merely an explanation of phenomena so that an informed decision can be taken. To inform this decision-making process will involve gathering, assessing and understanding multiple views (or truths) of how decisions are currently made and trialling existing frameworks for suitability as privacy decision-aids in practice.

Further, because privacy is a broad concept, to truly assess privacy, a thorough assessment of the privacy risks and potential impacts of any breach must be conducted. Such assessment however, must be done at an earlier stage, and as such, go beyond viewing technical security, access controls and privacy laws. Rather, it should encompass a more holistic assessment at a strategic level, considering the dataset as a whole. This will require for example, that the context in which the data was collected, controlled and created also be taken into account within any decision on whether a particular dataset poses a privacy risk if published.

Although there are existing frameworks that assess risk, none of these currently, in themselves, address privacy as part of such assessment sufficiently, particularly in the context of publishing data in open format. Therefore, this work will, in the first instance, seek to address this gap by creating a privacy-specific risk assessment framework for practitioners to follow in determining privacy risks associated with making data available as open data. To achieve this will involve conducting multiple qualitative case studies to gather data and trial suggested solution frameworks and, to that end, the following supporting research questions (RQ) will be asked:

RQ1 *How can existing risk and/or privacy decision frameworks or guidelines be adapted to create a privacy-specific decision-making framework that practitioners can adopt to support privacy decision making?*

RQ2 *How can contemporary legislation be incorporated into the privacy-specific assessment framework to practically support practitioners in privacy decision making?*

RQ3 *How can the privacy-specific assessment framework be adapted into a tool that can support any privacy decision making the organisation need to consider?*

1.2.1 Research Propositions

The underlying propositions (P) for answering the research questions are:

P1 *There are existing framework(s) that singularly or through amalgamation of concepts can be adapted to provide a practical foundation for determining privacy risks*;

P2 *The privacy assessment framework (PAF) developed as part of this study incorporates contemporary legislation and enables practitioners to systematically assess privacy risks in practice*;

P3 *A single tool or prototype can be created, based on the PAF, that can facilitate comprehensive privacy-specific decision making support.*

1.3 Approach

The approach for answering these questions will be to review privacy protection and, in particular, what legal provisions exist around privacy, so that these can be built into the final output of this work, the tool to support organisational privacy decision making.

Early work in this thesis, Chapters 4, 5 and 6, reflects the privacy law as it was up until May 2018. In May 2018, the General Data Protection Regulation (GDPR) came into force, resulting in a profoundly changed legal landscape around privacy. The provisions of GDPR were included because of the profound changes this regulation brought about in respect of privacy law. GDPR placed a number of much more stringent obligations on organisations and, as a result, changed the landscape of how privacy is, and will be, considered by organisations going forward (see Section 2.4.4). In recognition of this, GDPR was selected as the chosen *"contemporary legislation"* to incorporate into the framework created for answering RQ1 (Chapter 6). However, as the GDPR only came into force part way through this work, its provisions were incorporated as part of answering RQ2 in Chapter 7 and the final framework in Chapter 8.

As a result, it is contended that, in incorporating GDPR into this work, and the final framework (Privacy Throughout (PACT)), presented in this thesis, not only does the PACT framework practically support practitioners in privacy decision making ,incorporating contemporary legislation. It also facilitates multi-perspective privacy risk assessment, data lifecycle planning and enables organisations to demonstrate compliance to the relevant authorities. Thus, incorporating the provisions of GDPR altered the course of this work in a positive way to produce a truly versatile privacy decision making tool.

1.4 Report Structure

The remainder of this report is structured as follows. Chapter 2 reviews the existing literature and explains the terminology and concepts used in this body of work.

In Chapter 3, the methodology is presented and the chosen methodology, a case study, and reasoning for this choice outlined in Section 3.5.

This is followed by a series of scoping studies, carried out to provide direction and inform the research. The details of these studies and the findings from these, can be found in Chapter 4.

Following the results of the scoping studies, to answer RQ1 and test P1, a meta-model of the Contextual Integrity (CI) framework (Nissenbaum 2010) was created to demonstrate how an existing theoretical framework can be adapted into a formal, practical model in Chapter 5, from which a practical trial can be conducted, this model is outlined and discussed in Chapter 5.

The meta-model created in Chapter 5 is used as the basis for creating a prototype working framework that applies ContextuaL Integrity For Open Data in practice (CLIFOD). This framework is then trialled in a case study where CLIFOD was applied to a real-life scenario, using real data, in a practical setting, working in collaboration with a UK local authority (LA). The details of this study and the findings can be found in Chapter 6.

To answer RQ2 and test P2, in Chapter 7, CLIFOD was adapted to incorporate the General Data Protection Regulation (GDPR), that came into effect in May 2018, to create a Data Protection Impact Assessment (DPIA). This work formed part of a case study working in collaboration with a local charity ("the Charity") to implement GDPR and provide the Charity with a paper prototype for a standardised DPIA process (see Section 7.3.4).

The final case study, described in Chapter 8, answered RQ3 and P3 by creating an overarching prototype spreadsheet, PrivACy Throughout (PACT), that can be used to support all organisational privacy decision making for either new or existing projects, processes or systems. The thesis concludes with a summary of findings, discussion of the implications of the findings and future work in Chapter 9.

Chapter 2

Literature Review

2.1 Introduction

This chapter provides a review of the existing literature in relation to privacy and how organisations make decisions around privacy and privacy risks. It also provides a review of existing frameworks and guidelines pertinent to assessing privacy risk.

This review of the literature will begin by looking at privacy and what privacy means to different scholars before discussing privacy specifically in relation to data and public open data, the chosen area for testing how best to support effective privacy decision making. This area has been chosen because of the legal obligations placed on public bodies to release data in open format, thereby enhancing the risk of personal data being released unintentionally unless effective processes are or can be put in place, to assess the privacy risk before publication occurs.

This will be followed by a review of privacy protection from a practical, technical and legal perspective before considering how organisations currently make decisions around privacy and risk. The intention is that understanding these areas will help inform the creation of a privacy decision-making framework that will facilitate more streamlined privacy decision-making processes.

The rest of this chapter is organised as follows. Starting in Section 2.2, with a discussion about the meaning of privacy and how different perspectives of what privacy means contribute to the breadth of the definition of privacy and what this means in different contexts. This is followed by a section on privacy and open government and how open data and public open data affects privacy in Section 2.3. Next a review of privacy protection is provided in Section 2.4 and a discussion of some of the available interventions, tools and methods that organisations can deploy to protect data privacy. This includes a review of the practical (Section 2.4.1), technical (Section 2.4.2) and legal (Section 2.4.3) protection. This is followed by an explanation of the main provisions brought in by the General Data Protection Regulation (GDPR) which came into force partway through this work, in May

2018. This regulation introduced a number of profound changes to privacy law and, as a result, organisations face a number of new obligations as regards privacy. Therefore, an overview of these provisions has been included for completeness in Section 2.4.4.

In Section 2.5, an overview of organisational decision making is provided, followed by an exploration of risk in decision making in Section 2.6. This is followed in Section 2.7 by an overview of the relevant decision support frameworks and guidance documents available for practitioners to refer to for guidance in assessing privacy risk, before concluding the Chapter in Section 2.8.

2.2 Privacy

Privacy impacts many aspects of our lives and, while our expectations about how confidential information will remain if shared with friends, family, or even social media are likely to be met, this may not be the case with organisations or public bodies. At one time, privacy was mainly concerned with who divulged what and to whom. We now accept that government bodies need our personal information to conduct their business, and willingly provide details of our lives. Such details range from registering the birth of our children to providing details of our households on census days.

At the highest level, privacy concerns protection of self and our personal rights. This right to privacy is reflected in Article 12 of the Universal Declaration of Human Rights (The General Assembly of the United Nations 1948) and enshrined into law by the Council of Europe as part of the European Convention on human rights which states:

> *"Everyone has the right to respect for his private and family life, his home and his correspondence"* (Council of Europe 1950)

However, what constitutes an attack on honour or reputation? Perhaps a derogatory remark may be seen as an infringement of individual honour to some, while others may consider that same remark insignificant and take no offence. Thus, privacy has many facets, each person will perceive privacy in their own unique way and privacy tolerances vary from one person to the next. Therefore, privacy is difficult to define, as can be seen by how different scholars offer differing views on what privacy is. For example, privacy has been defined as:

A mental or psychological state meaning a condition state, being left alone or apart from others; (Weinstein 1971);

An extension of self or personhood (Moore 2015);

A freedom or a choice, for example, whether or not to participate (Parker 1974);

A power or control, this may include how much power we or others have over information pertaining to us i.e. *"the control we have over information about ourselves"*(Fried 1970); (Moore 2015);

An individual right or claim, this includes the right to self-determine what is communicated about us (Parker 1974); (Nissenbaum 2010);

A fluid concept this refers to the suggestion that privacy has blurred boundaries and in order to truly consider privacy these boundaries need to be considered. There are three types of privacy boundaries: disclosure, i.e. what we chose to disclose and what others might disclose; identity, i.e. our role as 'self' and how we interpret privacy the world; and temporality, i.e. our changing perceptions and time (Palen and Dourish 2003).

Irrespective of how privacy is defined, at an individual level, privacy affects all aspects of our lives, our beliefs, values, and cultural and societal norms. The privacy values of an individual are not rules or behaviours that are written down, rather they are individual to each person and influenced by the world and the culture that person lives in. This is based on the societal norms, personal values, and unwritten rules that individuals instinctively abide by (Nissenbaum 2010). How individuals negotiate life through their behaviours and thoughts involves, almost subconsciously, making decisions about these unwritten privacy rules.

Different professions view privacy from different perspectives, lawyers look at the legal framework (Shane 2015), while politicians may be more interested in public opinion. This may result in them attempting to balance perceived rights to privacy with safeguarding the nation against threats (e.g. to national security) (Nissenbaum 2010), or to appease public opinion, e.g. because of a public outcry (Michiel 2015). Similarly, on a technical level, statisticians may calculate the statistical probability of privacy being breached (Dwork 2006), whilst data miners will be more interested in what information can be gleaned from linking or mashing up the data (McAfee and Brynjolfsson 2012).

What most of these viewpoints have in common is that they are likely to only consider privacy specifically from their own perspective, without necessarily paying sufficient heed to how other perspectives might impact on this. Therefore, it can be argued that the multi-faceted aspects of privacy are not considered adequately in many scenarios. This may not be intentional but nevertheless, on occasion, everyone may fail to see the other perspective(s).

To account for these nuances, Westin (1966) talks about privacy based on the functions individuals perform in society, grouping these four categories; *"personal autonomy, emotional release, self-evaluation, and limited and protected communication"* (p. 1022). Solove looked at different aspects of privacy by dividing privacy into four broad groups that encompass 16 sub-areas for consideration: Invasions (covering intrusion and decisional interferences); information collection (covering surveillance and interrogation); information processing (encompasses aggregation; identification; insecurity; secondary use; and exclusion); and dissemination (covering breach of confidentiality; disclosure; exposure; increased accessibility; blackmail; appropriation and distortion) (Solove 2006). In devising this privacy taxonomy, what Solove sought to achieve was to create an overview of all the areas where a person's privacy might be compromised or breached. This therefore, includes a number of concepts that relate to an individual's physical environment as well as informational privacy.

This idea was expanded on by (Mulligan et al. 2016) who divided privacy into; *"five meta-dimensions of theory, protection, harm, provision and scope"*, sub-divided into 14 sub-dimensions looking at what or who is being protected (the object); against what or whom (target, from-whom); method of protection (provision) and the scope of that protection (boundaries). This approach is similar to a security threat analysis approach which could, for example, work very well within the technical design area and assist software design teams in aligning threat modelling with privacy (Faily et al. 2012), as can be illustrated by the work of Alshammari and Simpson (2018) who have aligned their privacy risk analysis framework with security threat modelling, to assess the potential harm caused by privacy threats and vulnerabilities. This is discussed further in Section 2.7.5.

2.2.1 Data Privacy

Looking more specifically at data privacy, while most of the frameworks discussed consider data privacy as one aspect of privacy, one framework, Nissenbaum's Contextual Integrity (CI), looks at privacy from a number of different perspectives but places more emphasis on data and how to assess the privacy risk of data flows in light of the surrounding context and how any change in the flow will impact on the context and therefore, the privacy risks (Nissenbaum 2010). CI asks that, in considering data privacy, not only must organisations consider the data itself, they must also take into account the context within which the data is processed, who handles the data and how it is transmitted in order to then place appropriate protection around the data (also see Section 2.7.7).

Thus, because this work seeks to establish is how organisations can be supported to make better, more informed decisions about data privacy and data sharing, the theory put forward here is that, CI provides an excellent candidate for use in this work. The holistic overview perspective of data privacy that CI offers makes this framework ideally suited for adaptation to create "a privacy-specific decision-making framework that practitioners can adopt to support privacy decision making" (RQ1, see Section 1.2).

The domain chosen as ideal for testing this theory, is the open data domain, in particular, the *public open data domain* (see Section 1.1.1). This area has been chosen as the starting point because of the legal requirements placed on public bodies to publish data in open format (see Section 2.3.1). In the open data publication domain the level of privacy afforded to individuals, is, at least in part, dependant upon how the organisation decides what constitutes personal data and when, how or whether a particular dataset will be shared or published. In particular, this work seeks to establish how public bodies make decisions around privacy risk as part of the publication process for open data publication (see Chapter 4). In order to understand these aspects however, it is necessary to first understand the motivation behind why public bodies seek to publish open data and to this end, open government, the main driver behind open data publishing, and open data will need explaining.

2.3 Privacy and Open Government

When a public body, or indeed any organisation, considers data, the two areas they are most likely to refer to will be first, the data itself and how this can be managed, stored and handled and second, any legal obligations or restrictions placed upon such handling of the data that might affect the individuals and/or the organisation. The individual, on the other hand, will just expect that their privacy will be preserved, particularly with public bodies who hold information about them. However, while the expectation of privacy may exist, the reality is that public bodies do not always get that aspect right.

> *"Ironically, one of the clearer threats to consumer privacy is the government's largely unchecked ability to collect your sensitive information without due process."*(Blackburn 2016)

Open Government has not been introduced under one set of regulations or a single act, rather, the path to open government has been introduced piecemeal through a variety of different enactments. The concept of Open Government, in its current format, became popular after President Obama released a Memorandum for the Heads of Executive Departments and Agencies in March 2009 encouraging increased transparency, collaboration and participation (see Figure 2.1). Since then other countries have also adopted the phrase, creating their own Open Government agendas including Australia, UK, New Zealand, Russia, China and the European Union (EC) (Wirtz and Birkmeyer 2015).

The political pressure to increase transparency between Government and it's citizens has further added fuel to the campaign for opening up access to government data (Obama 2009) to the extent that 70 countries have so far committed themselves to Open Government principles of widening citizen participation; encouraging more collaboration between citizens and their government and opening up access to government information (Open Government Partnership 2016).

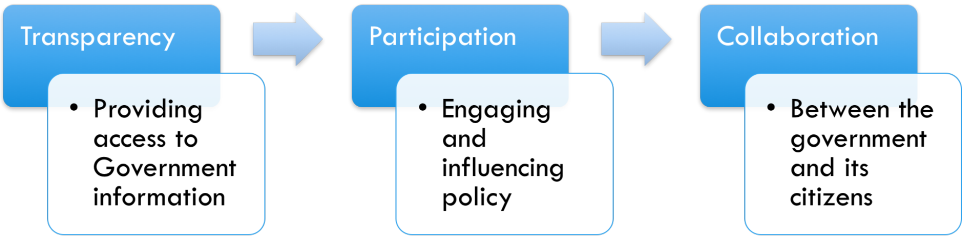

Transparency
• Providing access to Government information
Participation
• Engaging and influencing policy
Collaboration
• Between the government and its citizens

Figure 2.1: Open Government (based on principles in (Obama 2009))

2.3.1 Open Data

Open data is data that is freely available for anyone to download, with no restriction on re-use or redistribution (Open Data Institute 2016, Open Data Charter 2015) and public open data is open data published by a government body (Open Government Working Group 2007). This means that anyone who chooses to download the data are users. This may be an individual who is curious about a particular subject or who wants to gather evidence for a cause, for example, rising water levels in the local area. To investigate the user may access environmental data about water levels in an area released under the Infrastructure for spatial information in Europe (INSPIRE) regulations. Alternatively, a company may utilise open data to manipulate and amalgamate it for incorporation into their commercial product. For example, in Warsaw, Poland, one company has created an online platform that provides users with details of upcoming cultural events; while one entrepreneur in France has created a dataset that lists all cafe's in Paris who offer a cup of coffee for one Euro, all derived from public open data (Carrara et al. 2015). Thus, it is not only individuals who utilise public open data, many private sector companies are also users of public open data. There are many more examples of businesses generating value from public open data including big corporate businesses. For example, Tesco (one of the largest supermarket chains in the UK), IBM, Adobe, the Guardian newspaper and Yorkshire Water, all utilise public open data (Open Data Institute 2015).

Some private companies also collaborate and voluntarily share information with the public sector. For example, Uber shares data about trips with the city of Boston to help relieve congestion on the roads and manage urban growth (UrbanTide 2016); JP Morgan Chase supply banking authorities with data on their many subsidiaries; and Armajaro tracks its cocoa supplies and shares the information with both government and the local community to demonstrate their commitment to sustainability and meeting their labour and environmental pledges (Herzberg 2014).

Further, although there is no obligation on private companies to publish open data, some private companies choose to publish data either on their own or as part of a private/public collaboration. They may publish for research purposes, for example, Netflix made a large extract of their data available to the research community in 2007 (Narayanan and Shmatikov 2008) and the Enron dataset, was made available to purchase by Aspen at the direction of the Federal Energy Regulatory Commission (FERC) in 2003, and later released as open data for research purposes (Bartling 2015). Further, it has been speculated that this trend to release open data may spill over and become more commonplace with private organisations (Herzberg 2014). Thus, both private and public organisations publish open data. Hence, although the originator of open data may be either private or public, because of the legal obligation placed on public bodies, they are duty-bound to publish and therefore, they have an incentive to want to implement good practice in respect of privacy to avoid any potential for litigation by users whose data may have been compromised in the event of a breach.

Open Data and the UK Regulatory Framework

The sharing of government data ('public data') is central to the objectives of achieving Open Government. Allowing access to public body information conveys a government's commitment to transparency and encourages wider participation (Goda 2011). Moreover, as public bodies work on behalf of citizens, it has been argued that the public own the data collected and created (Shakespeare 2013). Consequently, it has been contended that citizens should be able to freely access that information in an open format (Fishenden and Thompson 2013). This sentiment is reflected in the Open Government Working Group's 8 principles of open government data which state that open data should be:

1. Complete - i.e. the raw data in unmodified format;
2. Primary - i.e. source data;
3. Timely;
4. Accessible;
5. Machine processable;
6. Non-discriminatory;
7. Non-proprietary; and
8. License free (although *"reasonable restrictions"* may be permitted) (Open Government Working Group 2007).

The requirement to publish public open data is not only a desirable objective, it is a legal obligation enacted through a series of statutes including: the Freedom of Information

Act 2000 (FOI); The Re-Use of Public Sector Information Regulations 2015 (ROPSIR); the Environmental Information Regulations 2004 and the Infrastructure for Spatial Information in the European Community regulations 2012 (INSPIRE).

In terms of this work, these obligations confirm that the choice of domain, *open data* (see Section 1.1.1), was a good choice of domain for testing the privacy-specific framework created as part of this work. Thus this domain can be selected not just because of the nature of open data, but also because of the legal obligation placed on public bodies to publish, although perhaps the domain should be more particularly chosen to be *public open data*.

FOI governs access to public sector information with some exceptions, for example, data classified as *"personal data"* under DPA will be exempt. Any individual may submit a request to a public body for information under FOI. Further, FOI also obliges public bodies to proactively publish data under a FOI publication scheme. The publication scheme was a mandatory requirement that prescribed, among other things, that information relating to policies, procedures, expenditure, and details of how decisions are made be published proactively (s. 19) (Information Commissioners Office 2015b).

ROPSIR requires public bodies to make 'raw' data available for re-use for secondary purposes. This means they must make the full dataset, including its metadata, available in machine readable and open format *"as far as is reasonably practicable"* (s. 11). However, there is no obligation on public bodies to adapt the raw data or extract from it in order to make it available for re-use (s. 6). Further, public bodies can impose license conditions on the re-use, providing this is non-discriminatory (s.13) and does not unreasonably restrict how the information is re-used (s. 12).

INSPIRE places a duty on public bodies to publish environmental geographical open data; and, public bodies are obliged to make environmental data available upon request under the Environmental Information Regulations 2004.

These obligations therefore place public bodies in a difficult position. On the one hand, they have lots of data which, for transparency and legal reasons they need to make available. On the other hand, they need to do so without compromising the privacy of the users and the data subjects whose information may be contained in the datasets to be published.

Public Open Data as a Commodity

Public open data is a highly valuable commodity, with estimates suggesting that globally, public open data is worth $3-5 trillion every year (Chui et al. 2014) and the European Union (EU)'s share of published Open Data was estimated at 55.3 billion EURO (Carrara et al. 2015) in 2015. Thus, it is clear that the open public data movement is going global and appears to be continuing to grow.

Research has shown that making public body data available in open format has indeed helped encourage innovation, bolstered citizenship, and promoted wider participation (Geiger and von Lucke 2012), making this a positive move. However, at the same time, concerns have been raised about whether, as a result of the release of open data, privacy is being compromised, despite claims to the contrary (Simpson 2011). This arguably, is a fair concern in view that several re-identification attacks on publicly released data have been successful (Henriksen-Bulmer and Jeary 2016, Ohm 2010), demonstrating that data breaches do occur, making this a very real threat. Further, such data breaches are not just a threat to privacy, they can have negative consequences for the users affected. For example, open data has been shown to create unforeseen personal security and privacy issues if personal details of individuals are released (Young and Verhulst 2016). Thus, failing to safeguard privacy and/or user confidentiality by inappropriately releasing or publishing personal data without consent can result in loss of public trust and potentially, personal damage claims from affected users.

According to the Open Knowledge Foundation (a UK not for profit agency), the commitment across the world to making public data available as open data continues to grow with 122 countries already providing some public open data (up from 97 countries in 2014) (Open Knowledge 2016). For example, the Open Data Charter, devised by G8 and endorsed by 11 governments including the UK, and 13 local government bodies, commits governments to making government data:

- Open by Default;
- Timely and Comprehensible;
- Accessible and Usable;
- Comparable and Interoperable' (Open Data Charter 2015).

In the UK, the Public Sector Transparency Board, set up to drive forward the UK Government's commitment to make Government data available in open format, have created 14 public data principles which incorporate most of the principles described in The Open Data Charter (Open Government Working Group 2007). However, neither set of principles make a full commitment to making the complete primary data available. The lack of commitment to completeness and making primary data available is unsurprising. Doing so would have serious privacy implications and it appears no-one has yet determined how to balance the data availability with privacy preservation of data subjects within that data. Simply referring to *”reasonable restrictions”* fails to sufficiently describe how to achieve this balance. Further, the increased transparency that open government promises through sharing data in open format, should also come with an increased responsibility to ensure the privacy of any individuals (”users”) whose information may be held within these datasets is maintained.

Thus, to facilitate open publication of this data and meet their legal obligations, public bodies need a robust privacy assessment framework that can help them ensure no personal data is published. To this end, they must consider how best to protect the data.

2.4 Privacy Protection

As part of protecting data privacy, organisations need to consider how they can protect users and the data subjects whose data may be held within a particular dataset. This requires a review of what protection, current or prospective, is afforded to the data and whether any additional protection and/or mitigation needs to be put in place in order to allow the data to be released or shared. This may be practical and/or physical protection, protection through technical controls, monitoring or intervention or, through legal adherence or compliance. Each will be dealt with in turn in Sections 2.4.1 to 2.4.4.

2.4.1 Practical Privacy Protection

There are a number of practical ways that organisations can protect privacy of their data other than from a merely legal perspective. This may include any combination of physical, administrative and/or technical controls. Physical controls may include camera surveillance, locks, fences etc. These can be used to monitor or restrict access to a physical location or limiting physical access to the room or building where the data is kept (Bauer et al. 2009). Administrative controls may include training staff on security and privacy matters in relation to data, and devising and implementing appropriate security and privacy policies within the organisation. ISO standards can assist in policy formulation and aligning these to the required purpose. For example ISO 9000 covers quality and customer satisfaction, ISO/IEC 20000 covers service management in Information Technology while BS ISO31000 and BS ISO/IEC29100 covers risk, security and privacy specifically. There are many frameworks and guides that can be used in devising appropriate policies which is outside the scope of this work and only a few are mentioned here for illustration purposes. However, one set of guidelines, the Privacy-Aware Government Data Release framework, is considered further as it deals specifically with policy development for releasing open data, this is discussed in Section 2.7.6.

2.4.2 Technical privacy protection

On a technical level there are many ways to protect data and the systems that house it, most of which tends to be geared towards security and how to protect the assets within an organisation, including data, rather than privacy. At the organisation level, this will most likely involve security and/or access controls starting with restricting unauthorised access to the IT infrastructure, networks and systems. There are three types of security controls:

1. Preventative - this may include for example, installing a firewall, an antivirus application and a requirement for all users to access systems using some form of authorisation and/or authentication;

2. Detective - this may include monitoring the network and/or systems and logs to discover any unusual activities in order to ensure early detection of any potential breach or threat e.g. using an intrusion detection system;

3. Corrective - this will include having appropriate mitigation strategies in place and evaluating and learning from incidents to prevent them re-occurring (Swire and Ahmad 2012).

At user or individual level, security controls may be placed on access to individual applications and/or systems. Access controls may be applied at network, system, device or data level depending on the organisation and sensitivity of the asset being protected. Further, to access the asset may require authentication (e.g. password or PIN) which requires the users to identify themselves and/or authenticate before being able to access or log in to systems (Sasse et al. 2001). This can be as fine-grained as requiring authentication to specific data within an application or system e.g. beyond access to the system, access can be further restricted (or layered) by role, name, data type etc., with each 'layer' requiring authorisation and/or authentication.

At system level, many solutions consider achieving privacy through security or system settings at a local system level. At the data level, the data itself can be encrypted prior to data transfer or data storage so that the data becomes unintelligible to anyone other than the sender and intended receiver. This can, for example, be done prior to electronic transmission so that only the intended recipient can access and read the data contained in a file (Piper and Murphy 2002). However, once data is shared beyond the individual system, alternative solutions need to be devised that can safeguard privacy in a larger, less controlled environment. The solutions derived for this purpose are many and varied and most solutions still consider privacy from the perspective of the attributes within the data and how these may be removed, distorted or obfuscated in some way. This can be achieved by anonymisation (Samarati 2001); pseudonymisation (Lablans et al. 2015, ICO 2012, Pfitzmann and Hansen 2010) or, through some form of partial or controlled technical release system (Dwork 2006, Machara et al. 2013). Thus, computer experts have had to create numerous ways to try to safeguard privacy and ensure that data protection is not breached.

2.4.3 Legal protection of privacy

From a legal perspective, most countries have some form of data protection legislation in place to safeguard personal data. For example, in Europe a centralised directive

(Directive 95/46/EC) governs data protection throughout the EU. This directive provides an overarching data protection framework for all EU member states, whereas in the United States (US), data protection regulations are de-centralised. Some regulations are enacted on a Federal level; these include the Fair Credit Reporting Act (FCRA), which covers privacy such as the prevention of identity theft; the Children's Online Privacy Protection Act 1998 (COPPA); and the Health Insurance Portability and Accountability Act 1996 (HIPAA), which stipulates what types of data is considered 'personal' in healthcare transactions. Other regulations are State-specific with most States also having some form of data protection in place (Swire and Ahmad 2012).

At the time of conducting the initial studies discussed in Chapters 4 to 7, a decision had not been reached as to whether the UK will implement GDPR in full. This position was clarified in May 2018, when the UK enacted an updated Data Protection Act 2018 (UKDPA), to bring GDPR provision into force in the UK after they leave the EU. However, for the purposes of the work conducted prior to this date, existing law at that time was assumed to prevail for the purposes of research conducted. Thus, to avoid confusion. For the purposes of the preceding work (Chapters 4 to 6), although the legal protection in place prior to GDPR was quite comprehensive in how data may be processed, the following issues remained:

1. Existing data protection laws dictate that personal information must be protected. This protection is limited to what the data controller and/or processor does with the data and so long as no 'identifying' information is published, that obligation has been met;

2. In establishing what is classified as personal data, DPA relies on assessment only of the sensitivity of the individual attributes within the data. It is therefore argued that data protection is one-dimensional because, once data has been redacted or anonymised and data subjects are no longer identifiable, under current law, the data is no longer considered personal or sensitive and thus, not subject to data protection (ICO 2012);

3. Data protection only requires the attributes within the dataset being considered, it does not require data linking or amalgamation to be considered as part of this assessment. This can result in any decisions around that data being considered in a 'silo' manner, which in turn can result in privacy being compromised. For example, when multiple datasets are combined or linked and used in new contexts, any 'silo-decision' made around one of the datasets will not have addressed the full range of potential privacy risks and thus could prejudice user privacy.

This is a concern as research has shown that anonymisation can be reversed (El Emam et al. 2011, Ohm 2010). Further, as datasets are aggregated or linked, the risk of re-identification increases (Henriksen-Bulmer and Jeary 2016). For those reasons, it has been argued that releasing public body data in open format was incompatible with data protection laws (Kulk 2012). This might however, be somewhat addressed by the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) which came into force in May 2018, the changes brought in by GDPR are discussed more in-depth in Section 2.4.4.

Prior to the introduction of GDPR, in the UK, Directive 95/46/EC was in force. This directive was enacted into UK law as the Data Protection Act 1998 (DPA). DPA governs what organisations and their employees may or may not do with personal data. DPA defines the actors or people who are the subjects of or involved in handling the data as:

- *"Data Subjects"* i.e. the individuals to whom the information pertains;
- *"Data Processors"* i.e. the people that process or handle the data on behalf of the data controller;
- *"Data Controllers"* i.e. the people responsible for decision making around the data.

Together, data processors and controllers must ensure that the eight data protection principles are upheld where personal data is processed. To that end they must ensure:

1. *Lawful* - any processing is only conducted for lawful purposes;
2. *Limited* - data is only used for the specified purpose(s);
3. *Adequate* - used in a relevant, adequate and non-excessive manner;
4. *Accurate* - data is correct and accurate;
5. *Minimal retention* - data is not retained beyond what is absolutely necessary;
6. *Respectful* - data is handled in adherence to people's data protection rights;
7. *Secure* - data is stored and maintained securely;
8. *Safe* - ensure data is not transferred outside the European Union (DPA).

To clarify how this applies however, some clarification on definitions may prove useful.

Data Processing

The handling or working with data is referred to as *"data processing"* under DPA, with *"processing"* referring to any handling of the data, be that collecting, manipulating, storing, processing or any other handling of the data (DPA, s. 1(1)). Therefore, the handling of the data will be referred to as data processing.

When storing data, according to relational database theory, data is held in tables (Abiteboul et al. 1995), an example is provided in Table 2.1 for illustrative purposes. Tables contain rows (entity sets), columns and fields (entity or tuple). Each column will hold a specific type of information with common characteristics such as name, address or gender, with each column holding data with similar characteristics. For example, the name column might contain all the first names of the subjects in the database e.g. Alice, Bob and Eve, these are known as attributes. The rows contain a set of data from each column relating to a record, e.g. a row might contain; Alice (name), A Street (address) and female (gender), these will all be in the same row or 'entity set'. The fields (entities or tuples) then contain the specific information pertaining to each column (the individual attributes) (Chen 1976).

Row (Entity Set)	**Name**	**Address**	**Gender**
1	Alice	A Street	Female
2	Bob	B Street	Male
3	Eve	E Street	Female

Table 2.1: Example Database Table

Personal Data

Under DPA, personal data is defined as; *"data which relate to a living individual who can be identified (a) from those data, or; (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller"* (DPA s. 1). Thus, to ensure compliance, the individual elements of data within a dataset must be considered (the 'attributes') in order to determine whether or not they should be classed as personal data. Presumably therefore, if we remove or obfuscate any personal attributes, then DPA is satisfied and we need not worry.

In order to ascertain whether data is personal, data can be classified according to how sensitive or identifying it may. To this end, attributes can be broken down into four types;

- *Identifiers*, i.e. data that can directly identify an individual, such as their name, national insurance number or date of birth;
- *Quasi-identifiers*, i.e. data that is not directly identifying but likely to be if linked, e.g. age or gender (Thomson et al. 2005);
- *Sensitive attributes*, i.e. individual specific data that could aid in identifying an individual, such as ethnic origin, religious beliefs, disease or salary (Fung et al. 2010);
- *Non-sensitive attributes*, i.e. non-identifying, even if linked.

However, the definition of what constitutes personal data does vary between countries. For example, in the US, when talking about personal data, this is referred to as personal identifiable information (PII) and defined as:

1. Any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and

2. any other information that is either linked or linkable to an individual, including medical, financial, educational, and employment information. (NIST 2010)

It means that, in general, in the US, information that falls into either of the first two categories above ('identifiers' and 'quasi-identifiers') are likely to be classed as PII while, information classified as sensitive will not be classed as PII. However, this is not the case in Europe where the definition of what is personal data is much broader. In Europe, personal data is 'ANY' data that can identify a person and therefore this will include sensitive data that is outside the scope of the PII rules.

2.4.4 GDPR

The introduction of GDPR brought about considerable change in privacy law and introduced a number of additional and new obligations on organisations to make sure their data processing adequately safeguard the data subjects. Because GDPR is a regulation, it takes effect immediately in all member states within the EU from the effective date. This means that, in contrast to a directive, which must be enacted by each member state within the EU, the regulation is legally binding in its entirety from the effective date across all member states, no enactment at national or individual Country level is necessary (Europa 2018). In view of Brexit there was some initial confusion as to whether or not the UK would enact the GDPR regulation into UK law after leaving the EU. This has now been resolved and the UK Government have enacted an updated Data Protection Act, UKDPA, devised to implement GDPR provisions into UK law (UK Parliament 2018). The provisions brought in by GDPR afford data subjects extensive rights including the right to be forgotten (Article 17) and stricter rules have been placed around consent. These state that consent must be freely given, explicit (Article 7) and provided by a *"clear affirmative act"* (Recital 32) (European Parliament and the Council of Europe 2016).

Because GDPR imposes a number of new and/or tightened provisions that organisations must adhere to, including a requirement to protect and safeguard data relating to individuals (i.e. personal and sensitive data), these provisions will have to be incorporated in this work. However, it should be noted that because, GDPR came into force part way through this PhD, early work conducted was based on the legislation in force at that time, i.e. DPA (Chapters 4, 5 and 6). Later work however, looks at how GDPR affects organisations and incorporates this into this work.

GDPR Principles

Chapter 2 of GDPR details the principles upon which the Regulation is based. The rules relating to data processing are laid out in Article 5 which sets out the following 6 overarching Principles:

Principle 1: Lawfulness, fairness and transparency *Lawful*: Organisations should determine and define under what legal basis they are processing the data (Article 6 and 9); *Fair*: Data should be processed and used fairly, with data subjects interest in mind; *Transparency*: Data subject(s) must be kept informed about what data processing will be done, and why (Article 12). Information should also be provided about how the data will be used (Articles 13 and 14). Further, this information must be relayed to data subjects in plain, easy to understand language (Recital 39);

Principle 2: Purpose Limitation Before processing can be done, the specific purpose for processing must be established and communicated to the data subject. Further, no processing beyond that original stated purpose is permitted without prior informed consent from the data subject (Article 7, Recital 32). If further processing is required not only must the data subject provide informed consent to such further processing, the legal basis for the proposed or planned further processing must be established as well, it cannot be assumed that the same legal basis for processing will apply (Recital 50);

Principle 3: Data Minimisation Data collected should be adequate, relevant and not excessive. This means you should collect minimum data i.e. only collect the minimum data that is needed for your specified purpose (Article 25(2));

Principle 4: Accuracy Keep data up to date and accurate;

Principle 5: Storage Limitation Only keep data for as long as absolutely necessary (Recital 39); anonymise, pseudonymize and delete data as soon as it is no longer needed for the original purpose (Article 25); and securely delete and/or destroy data no longer needed;

Principle 6: Integrity and confidentiality *Integrity*: Safeguard the accuracy and completeness of the data; *Confidentiality*: Process and store the data securely ensuring data is protected from harm, unauthorised or unlawful disclosure, use, modification, damage or access (Recital 39). Ensure all staff are trained in how to safely & securely handle and process data. *Security*: In order to safeguard and protect the data, organisations must ensure they have appropriate security measures in place both organisationally and technically. This should include polices, processes and technical measures, involve assessing privacy risks and implementing suitable mitigation strategies, taking into account the risk likelihood, severity and the purpose, context nature, scope and cost of implementation of any security measures (Article 32).

From the perspective of the organisation, they, as data controller, must establish the legal basis upon which processing is carried out up front (Articles 6 and 9) and ensure appropriate technical and procedural measures are put in place to safeguard the privacy of the data subjects. Further, they must be able to demonstrate compliance (Article 24) and adopt a *”privacy first”* policy (*”privacy by design and default”*, Article 25). GDPR does provide some guidance for how organisations may demonstrate compliance by asking that they maintain a record of processing activities (Article 30) and ensure adequate security measures are in place to protect the data (Article 32). Finally, GDPR places and obligation on organisations to conduct Data Protection Impact Assessments (DPIA) for any processing activities classed as *”high risk”* (Recital 90, Article 35). In the following sections, the main points that organisations need to be aware of are outlined.

For the purpose of this work, because CI has considers privacy from a holistic viewpoint, GDPR is explored using the CI perspective to demonstrate how the GDPR principles have been described according to how they relate to the main contextual integrity perspectives: *attributes; actors; and transmission principles*.

GDPR *Attributes*

Article 4, defines data as: *”any structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis”* (European Parliament and the Council of Europe 2016). Thus, any data that includes personally identifying information will fall under the protection of GDPR which, in terms of this work and CI, will therefore refer to the attributes, i.e. the individual data items (see Chapter 5, Section 5.4.1).

Record of processing activities

The rules relating to the organisations (as data controller and/or data processor) and their obligations under GDPR are laid out in Chapter 4 of GDPR. These include the obligation to implement data protection by design and default (Article 25, see also Section 2.4.4), report any data breach within 72 hours of first finding out that a breach has, or may have, occurred (Article, 33(1)), and a requirement to keep a record of *”processing activities”* (Article 30). This should include records of:

People i.e. actors which in terms of CI is the *data; sender, receiver and subject as detailed above*:

- Responsible Person i.e. noting who the person designated by the data controller as responsible for overseeing data processing and, if applicable, the designated Data Protection Officer (DPO) are;

- Data Processors i.e. sub-processors referring to any third party individual or organisation who will be processing the data on behalf of the data controller;
- Data Subject and Categories: a description of the data subjects and data categories being processed (e.g. personally identifying, sensitive etc.); and
- Stakeholders: details of any third-party stakeholders with whom the data will be shared.

Data : i.e. the attributes as described in Section 2.4.4. This should record the:

- Legal Basis for Processing: the lawful basis for processing the data, there are six acceptable lawful reasons for processing personal data (Article 6) and, if processing involves special category data (such as health data or sexual orientation), a secondary, special category justification will also need to be noted for processing personal data (Article 9);
- category: data should be categorised (see above) and justification should be noted for collecting each item of data;
- storage: the length of time data is stored and time limits for each data category;
- mitigation: a description of risk mitigation (security) measures employed both organisationally and technically.

GDPR *Actors*

Under GDPR the actors are similar to the actors under DPA with some additional duties and roles added. For example, a new role of Data Protection Officer (DPO) has been created to oversee that GDPR is implemented and adhered to within the organisation. This role is advisory for smaller organisations but compulsory for public bodies and any organisation whose *”core activity”* includes the processing large scale and/or quantities of personal data (Article 37). The other main actors and roles under GDPR can be described in terms of CI as follows:

Data Sender(s) The data sender(s) are the *data controller* i.e. the person or organisation legally responsible for processing the data; *data processor* i.e. any third-party person or organisation who processes the data on behalf of the data controller (Article 28) and/or *joint data controller* i.e. any person or organisation who processes data collaboratively with the data controller (Article 26);

Data Subject the person who is the subject of the data (see 2.4.3) who, under GDPR, have a number of rights:

Right to be informed Data subjects have the right to be informed of how their data is used in; *"clear and plain language"* (Article 12, Recital 58);

Right of Access Individuals have a right to be informed of what processing is being done with their data and to receive a copy of information held about them (Article 15);

Right to Rectification The data subject can request that any inaccurate data about them is corrected and rectified (Article 16);

Right of Deletion (a.k.a. 'the right to be Forgotten') the data subject has the right to ask for their data to be deleted from the organisation's files (Article 17);

Right to Restrict processing Where the data subject is disputing the lawfulness of the processing; contesting the decision of the data controller; or has invoked one of the other rights that have either been refused, he/she can request processing is restricted until the matter is resolved. (Article 18);

Right to Notification Be notified regarding any deletion, restriction or rectification requests (Article 19);

Right to Data portability i.e. receive a copy of his/her data or have this transmitted to a third party of his/her choice (Article 20); and

Right to Object The right to object (Article 21).

Data Receiver This is the recipient which may be an internal or external stakeholder who receives and processes the data i.e. the data processor and/or joint controller or, where consent has been obtained, this may be an authorised third party.

Consent

The rules regarding consent are laid out in Article 7 of GDPR. This states that, where the legal basis for processing is based on consent, the organisation must be able to demonstrate that the data subject has given informed consent (Article 7(1)). Further, such consent must be *"freely given"* (Article 7(4)). However, where an imbalance exists between the data controller and the data subject, it is possible that consent may be presumed not to have been given freely (Recital 42). Moreover, consent cannot be considered freely given where providing a service is dependent on the consent unless it is necessary for performing the contract (Article 7(4), Recital 43).

GDPR *transmission principles*

How the data is processed and shared is covered in CI under the *transmission principles*. In terms of GDPR, the transmission principles need to be considered for both existing and new processing. The way the regulation deals with this is by asking data controllers to establish a lawful basis upon which any data processing is done. There are 6 'permitted' lawful basis' as follows:

Consent data Subject has given informed consent to the processing;

Contractual data subject is party to a contract where processing is necessary for the establishing of or performance of the contract;

Legal Obligation processing is necessary for compliance with a legal obligation placed upon the data controller;

Protection of Vital Interests processing is necessary in order to protect the vital interests of the data subject or of another natural person;

Public Interest processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

Legitimate Interests processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child (Article 6).

Where certain sensitive (aka *"special category"*) data is processed, such data is also classed as personal under GDPR (Article 9). Therefore, processing of such sensitive data must also be legally justified and a lawful basis for the processing recorded (Article 9(1)). There are ten 'permitted' legal basis upon which *"special category"* data may be processed:

Consent data Subject has given informed consent to the processing;

Necessary processing is necessary for meeting a legal obligation in regards to employment, social security or protection;

Protection processing is necessary in order to protect the vital interests of the data subject or of another natural person where data subject not legally capable of providing consent;

Member processing processing carried out in relation to members or former members for legitimate activities in relation to legitimate activities of the association, foundation or not-for-profit organisation;

Publicly available data processing relates to personal data that the data subject has themselves already made public;

Legal Claim processing is necessary for a establishment, exercise or defence of a legal claim;

Public Interest processing is necessary in the substantial public interest;

Preventative processing is necessary for establishing occupational working capacity of the data subject or for medical reasons to provide care, diagnose, prevention or treatment;

Public Health processing necessary in the interest of public health;

Research processing necessary for archiving purposes in the public interest or for statistical or historical research purposes.

Data Breach & Notification

Another part of the CI transmission principles is to review where these have been/might breached which forms part of the privacy assessment. Article 32 requires that data is kept and processed securely. Where a breach occurs, or is suspected, Article 33 places an obligation on organisations to notify the relevant authorities (i.e. the Information Commissioner's Office (ICO) in the UK) within 72 hours of first finding out that a breach has, or may have, occurred (Article, 33(1)). This means that notification in many instances, will have to be made before confirmation that there has been an actual breach has been established (Recital 87). If the breach is deemed high risk to the data subjects and likely to affect their rights and freedoms, the data subjects should also be notified, without delay, that the breach has occurred (Article 34).

Privacy by Design and Default

Article 25 places an obligation on organisations to implement privacy by design and by default. As discussed in Section 2.7.4, privacy by design refers to organisations being required to carry out privacy risk and impact assessments. These should be conducted as standard to establish what potential privacy risks might be associated with the data processing carried out by the organisation including consideration of the likelihood and potential severity of any such risk materialising. Under GDPR however, the PbD principles have been translated into a more specific Data Protection Impact Assessment (DPIA), which must be conducted for any new process, system or project (Article 35), thereby making PbD obligatory, rather than advisory. Further, the organisation is required to put in place appropriate technical and procedural protection to safeguard the processing of the data (Article 25(1)).

Privacy by default, places an obligation on the organisation to also must ensure they only collect data that is necessary for the agreed purposes. This means making sure that only appropriate data is collected and each piece of data collected is justifiable. Further, implementing privacy by default also entails ensuring that technical and procedural protection is in place to safeguard the data; how the data is stored; the length of time the data is stored; and who can access the data (Articles 25(2) & 32).

Data Protection Impact Assessment (DPIA)

Any processing of personal data that is likely to pose a high risk to the rights and freedoms of the data subject must be assessed to determine the likely risks associated with such processing under Article 35. This places an obligation on organisations to assess risks, not from an organisational perspective, but from the perspective of the data subject (the individual). A list of when a DPIA is required and/or advisable can be found in Appendix D, Section D.9.2.

Ultimately, irrespective of which regulation or law applies, privacy must be integrated into organisational decision making and thus built into corporate practice (Bamberger and Mulligan 2015). This requires privacy to be considered beyond the level of attributes within the data itself and whether or not a particular attribute within the data is sensitive and thus must be pseudoymised or anonymised. Lederer et al. (2003) argue that to truly consider privacy, a system should incorporate the; *"system properties, actor relations and information types"*, as these all form part of the privacy space. This will require a look at how organisations make decisions around data, not just from a practical, technical or legal protection perspective, but strategically and in a more holistic manner.

2.5 Organisations, Data and Decision-making

Organisations, whether private or public, collect information. Some personal information is provided voluntarily by users or citizens (hereafter referred to collectively as 'users'). For example, when purchasing a product a user will provide the company with details of their address etc. for delivery or, when applying for insurance or credit, a user will disclose more in-depth information including perhaps, their date of birth, their family relations and a copy of their passport. In the case of public bodies, they collect 'official' information about users, i.e. information that is collected and used by government bodies about users for an official purpose. For instance, citizens are required to inform authorities when they get married, have a child or a loved one has passed away. They also provide details for tax, electoral roll, applications for grants or financial assistance, and housing to name but a few.

Westin was the first to observe the right to control personal information rather than autonomy or dignity. Further, he recognised that the right to privacy is context dependent (Westin 1966). Similarly, the value of data is content dependent. Once data has been analysed and interpreted, it turns into information, which in turn may be interpreted and decisions or wisdom, derived from it (Ackoff 1989). When organisations collect information about users, they do not hold each piece of information as a separate stand-alone entity, they collate and/or process this information in some way to make it useful to their purpose, thus, they turn the information into data.

Data, is a collection or *"set of data"*, that may be stored electronically (Oxford University Press, 2017). The main purpose of data is to capture activities or events and therefore, most data is historical (unless it is used for forecasting or illustration purposes Liew (2007)).

Information, on the other hand, is the; *"knowledge communicated concerning some particular fact, subject, or event; that of which one is apprised or told"*(Oxford University Press 2017), i.e. it is the meaning that may be derived from the data either through processing, decisions taken or analysis (Liew 2013).

Businesses derive all manner of important detail from the information collected to serve their particular purpose. This may be to post the product ordered, process the credit application or, it may be required for government bodies to carry out their functions. However, the data may also be used to discover more about the individuals and their preferences. For example, businesses may use and/or generate data which tracks users online movements to establish what they like and record purchasing habits for marketing purposes or public bodies may publish open data to meet their transparency obligations as part of Open Government. Information however, may be gleaned from the processing or analysing of the original data, with or without the consent or knowledge of the users who supplied the data.

That is not to say however, that all processing of data is negative, data promises many exciting opportunities for research, data analytics and business gaining competitive advantage using the power of that data (McAfee and Brynjolfsson 2012). Data analytics can provide powerful insight and inform business decisions. It can also be used to help prevent fraud, manage staff, determine risk exposure, or inform how best to operationalise business strategy (Acito and Khatri 2014).

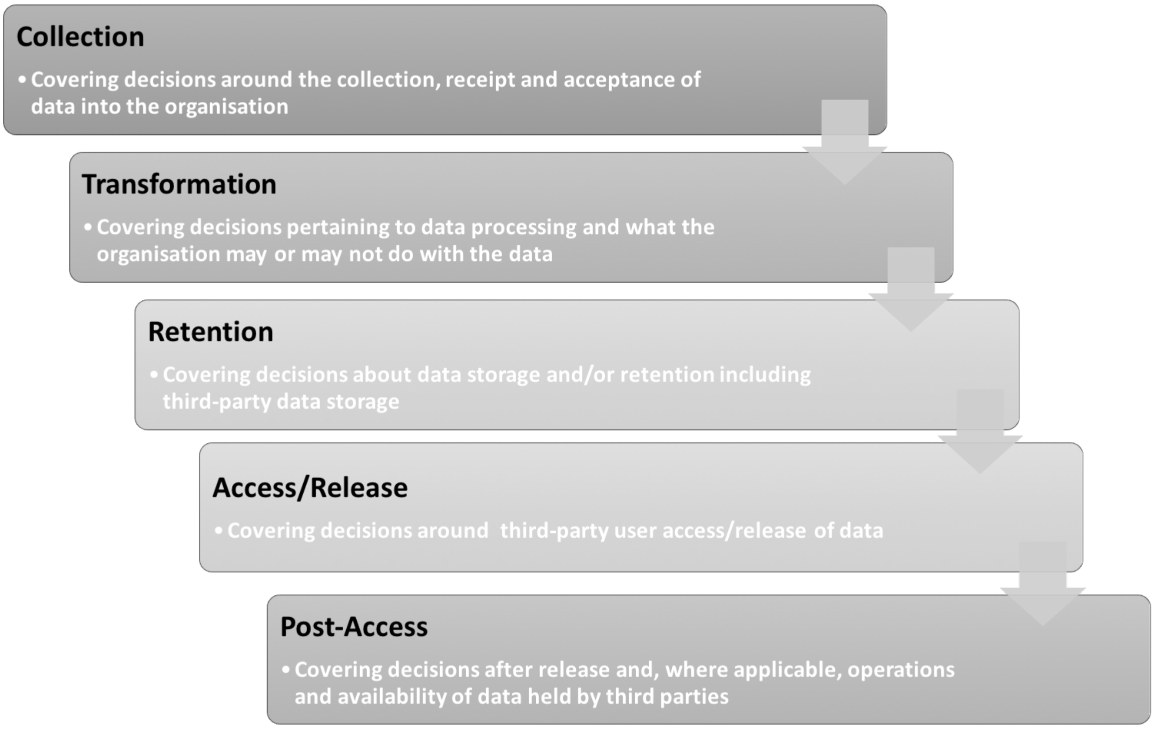
Once organisations have the data, they need to make decisions about what to do with the data; how to handle, store, process and use the data and whether to retain and/or share the data. In answering some of these questions, it is helpful to first break down where in the data lifecycle the data sits in order to break down the decision-making into more manageable chunks because different decisions will need to be made at different stages in the data lifecycle.

The data lifecycle is a 5-stage data lifecycle model developed by (Altman et al. 2015). This model was created to categorise suitable security measures that can be employed by a public body to preserve privacy. Data release is broken into the following lifecycle stages:

- Collection - this is the collection, receipt and acceptance of data into the organisation;
- Transformation - this is the data processing stage where the organisation will work with the data;
- Retention - storing of the data, including third-party data storage;
- Access/Release - third-party user access/release to the data;
- Post-Access - operations and availability of data held by third parties. (Altman et al. 2015).

However, Altman's framework can equally be applied to the decision-making process to provide a useful division of areas within which decisions need to take place regarding data management and/or release. For example, early in the data lifecycle, when data is being collected, the decisions that need to be made will be focused on informing the data subject about what the purpose of the data collection is, how the data will be used and ensuring consent has been obtained. Later in the life cycle however, the decisions that need to be made may focus on whether or not to destroy or retain the data.

To illustrate, Figure 8.6 shows how this model can be adapted and applied to the decision-making process to provide a practical division of areas within which decisions need to take place in regards to data management and/or release.



Collection
- Covering decisions around the collection, receipt and acceptance of data into the organisation

Transformation
- Covering decisions pertaining to data processing and what the organisation may or may not do with the data

Retention
- Covering decisions about data storage and/or retention including third-party data storage

Access/Release
- Covering decisions around third-party user access/release of data

Post-Access
- Covering decisions after release and, where applicable, operations and availability of data held by third parties

Figure 2.2: Adaptation of (Altman et al. 2015)'s data lifecycle model for decision-making

Looking in particular at the area of focus in the first case study, open data publishing, decision-makers will be looking to publish existing or historic data and determine whether each historic and/or existing dataset can be published as open data. Therefore, in this area, decision-making will focus on data suitability during the access/release and/or post-release stages of the data lifecycle. This in turn, requires an understanding of how businesses make decisions.

2.5.1 Decision Theory

In decision theory, decisions are made either by *"economic man"* who makes rational decisions based on accurate information or *"organisation man"* whose rationality is bounded by the decision maker's intellectual and/or knowledge limitations and external societal or social constraints (Simon 1955). Economic man has three properties:

1. He is rational - meaning he can maximise his choices and can weakly order his preferences to achieve what he requires to do this;
2. He is fully informed - he knows both what the options are and the outcomes of each option;
3. He is boundlessly sensitive - meaning the options available are differentiable and continuous and therefore calculable (Edwards 1954).

Rational man makes riskless choices, he knows all the variations and can calculate likely outcomes using indifference curves. However, even in economics, the idea of riskless choices are no longer considered realistic. Rather, the concept of risky choices were introduced for economic man, where risks and options are calculated in terms of probability of occurrence and expected value (Edwards 1954). Another decision theory is game theory, based on mathematical programming and statistics. This offers little practical help in developing strategies, but it does offer rules about how to choose among a set of options (Edwards 1954). Today's economic man therefore, may calculate probabilities of occurrence or indeed, he may calculate financial risk or exposure using formulas such as Net Present Value (NPV), Rate of Return (ROR), Benefit-Cost Ration (BCR) or Total Cost of Ownership (TCO) (Bhushan and Rai 2007). However, not all risks are tangible and adding a value to these requires a little more ingenuity in adding some form of economic value (see Section 2.6.1). This is where the decision style of *"organisation man"* comes to the fore.

In contrast to *"economic man"*, *"organisation man"* does not know everything, rather he is assumed to consider;

1. *Context and Unknowns* - he determines the context and objective to be decided upon and understands that he will not be in possession of all the facts or detailed knowledge of the consequences of actions taken;

2. *Search* - he must therefore process what he knows and consider alternative solutions to the problem;
3. *Decide and act* - he must determine the best course of action based on the available alternative solutions (Simon 1955 1979).

Further, *"organisation man"* will, in going through the decision-making process, during the search phase, stop at a 'satisfactory' solution rather than continue searching for an *"optimum"* solution (Simon 1955). This has also proved to be the case in studies of management decision-making in business (Fleming 1966, Cowling and Sugden 1998). Accordingly, organisational decision-making consists of; *"a moment in an ongoing process of evaluating alternatives for meeting an objective, at which expectations about a particular course of action impel the decision maker to select that course of action most likely to result in attaining the objective"* (Harrison and Pelletier 2000). This, in turn is likely to involve *"some random elements...[and] many decisions are [therefore] subject to risk"* (Galanc et al. 2016).

As a result, *"organisation man"* will make decisions based on the information available to him. This may not be perfect information but it is enough for him to act and make a decision. To aid in this decision-making process, *"organisation man"* may either conduct or request some form of measure or calculation be carried out on the likely outcomes of a particular choice. To this end, the decision-making process may well take into account the risk or the *"risk-reward factor"* (Harrison and Pelletier 1998), i.e. considering what level of risk and/or return a particular decision is likely to result in because; *"between calculated risk and reckless decision-making lies the dividing line between profit and loss"*(Duhigg 2005). Relating this to the likely users for the output of this work, the practitioner or users most likely to use the framework will be *"organisation man"*, as they will be unlikely to have the option to make riskless choices, nor will they have all facts available to them.

2.6 Risk

In today's business world, risk is a popular method for trying to define the exposure, cost or potential impact of a decision, or alternate solutions to a problem with an organisation (Simon 1955). Risk is an attempt to define the uncertainty in more practical terms. Indeed, it has been contended that risk *"is inseparable from decision-making"* (Galanc et al. 2016). Thus, this work will follow this example and adopt a risk based approach using the decision theory of *"organisation man"*.

Using risk to help inform and guide decision making is common practice in most Countries. To this end, in the US, the National Institute of Standards and Technology (NIST), have issued general guidelines pertaining to risk (NIST 2012). Similarly, the International Organisation for Standardisation (ISO) have produced a series of risk guidelines which are recognised as worldwide standards for both private and public bodies for assessing risk (e.g. ISO 31000:2009), that have been adopted and recognised internationally in 160 Countries (Swire and Ahmad 2012). These are described briefly in the following sections.

NIST

NIST have, as part of their general risk guidelines, devised a standard information security risk assessment process for organisations to follow in making risk-based decisions, this process is based on a formal lifecycle that includes;

1. Framing the risk;
2. Assessing the risk;
3. Responding to the risk;
4. Monitoring the risk. (NIST 2012)

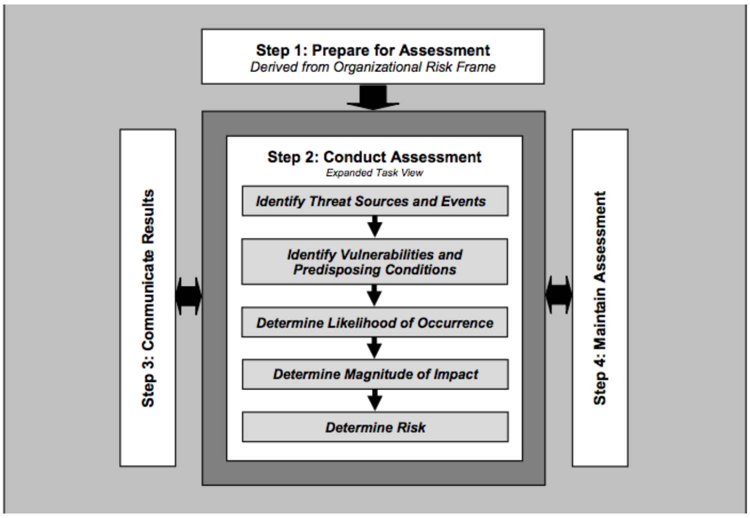

Step 1: Prepare for Assessment
Derived from Organizational Risk Frame
Step 2: Conduct Assessment
Expanded Task View
Identify Threat Sources and Events
Identify Vulnerabilities and Predisposing Conditions
Determine Likelihood of Occurrence
Determine Magnitude of Impact
Determine Risk
Step 3: Communicate Results
Step 4: Maintain Assessment

Figure 2.3: NIST Risk Assessment Process (NIST 2012)

The guidelines go on to recommend a more detailed risk assessment process which is outlined in Figure 2.3.

The idea of adopting a risk based approach to decision making is widely accepted and recommended in the public sector around the world including: the US (Office of Management and Budget 2013); Europe (FERMA 2003); Canada ((Province of British Columbia 2012)); and South Africa (National Treasury 2010). The ISO standards are the recommended guidelines within Europe, Canada and South Africa.

ISO 31000

BS ISO 31000:2009 was devised to provide generic standardisation of risk management principles and provide guidelines that would span all industry sectors and represent risk management *"best practice"* (Hall 2011) and was adopted and implemented in the UK by the British Standards (BS) office in 2010 (BS ISO 31000 2009).

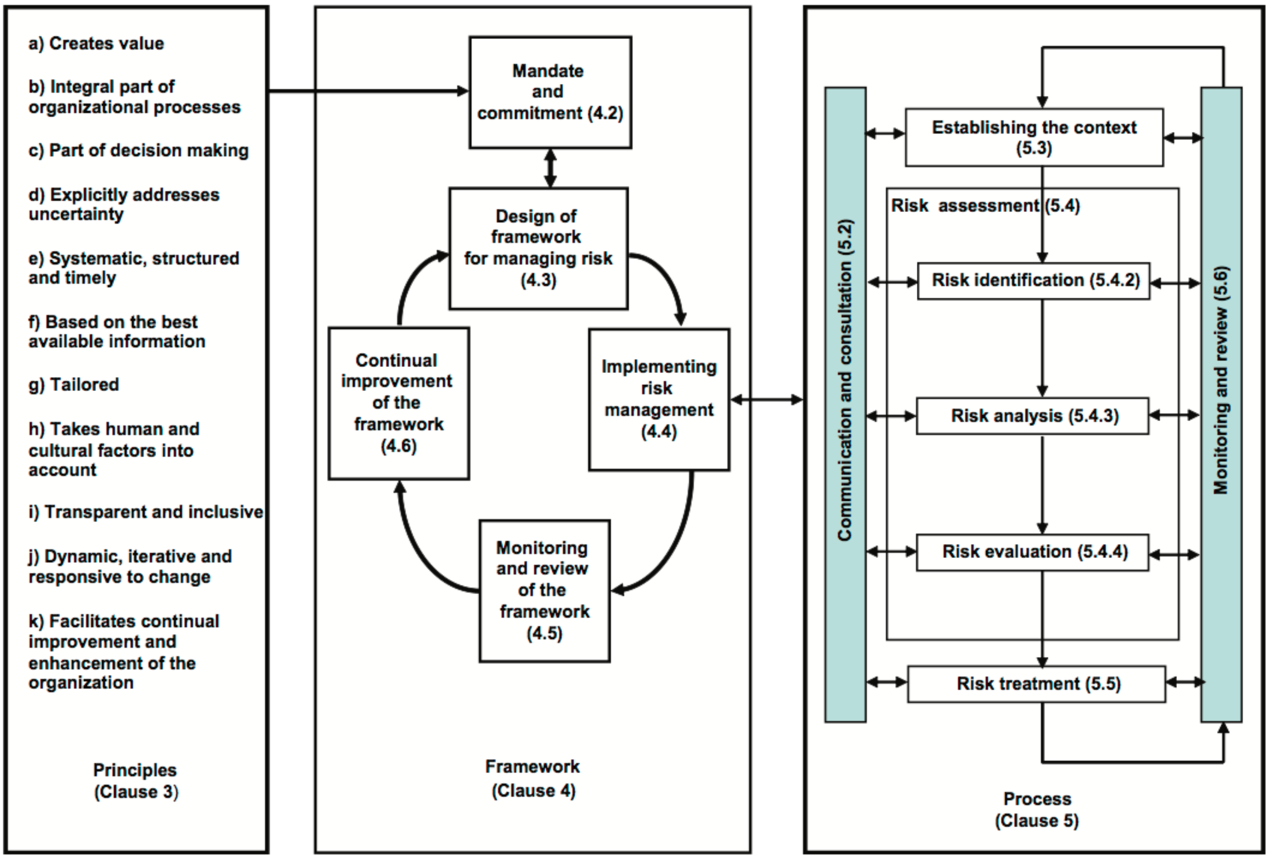

a) Creates value
b) Integral part of organizational processes
c) Part of decision making
d) Explicitly addresses uncertainty
e) Systematic, structured and timely
f) Based on the best available information
g) Tailored
h) Takes human and cultural factors into account
i) Transparent and inclusive
j) Dynamic, iterative and responsive to change
k) Facilitates continual improvement and enhancement of the organization
Principles (Clause 3)
Mandate and commitment (4.2)
Design of framework for managing risk (4.3)
Implementing risk management (4.4)
Monitoring and review of the framework (4.5)
Continual improvement of the framework (4.6)
Framework (Clause 4)
Communication and consultation (5.2)
Establishing the context (5.3)
Risk assessment (5.4)
Risk identification (5.4.2)
Risk analysis (5.4.3)
Risk evaluation (5.4.4)
Risk treatment (5.5)
Monitoring and review (5.6)
Process (Clause 5)

Figure 2.4: ISO31000:2009 - Risk Management principles, framework and process

The ISO contends that defining, managing and assessing risk establishes *"a reliable basis for decision making and planning"* (BS ISO 31000 2009). Risk is defined as; *"the effect of uncertainty on [an] "objective" where "uncertainty" is the state, even partial, of deficiency of information related to, understanding or knowledge of, an event its consequence, or likelihood"* (BS ISO 31000 2009).

This standard consists of ten principles, a risk management framework and a risk management process as depicted in Figure 2.4.

Figure 2.4 shows that the risk principles are concerned with assigning value to uncertainty in order to form part of the decision-making process taking into account available information and external factors that may affect the risk. This is similar to Simon's bounded rationality theory principles (Simon 1955 1979). However, this link is perhaps more evident when Simon's bounded rationality theory is compared to NIST and the five steps in the

ISO 31000:2009 risk management process as shown in Table 2.2.

BS ISO 31000:2009	**(Simon 1955)**	**(NIST 2012)**
Context the process starts with establishing the objectives within the given context	*Context and Unknown*	Identify threats and vulnerabilities (Assess)
Risk & Analysis decision-makers are asked to identify and analyse the risks	*Search*	Determine likelihood and impact (Monitor)
Evaluation and treatment completing with an evaluation and treatment	*Decide and Act*	Determine risk (Respond)

Table 2.2: ISO31000:2009 vs (Simon 1955) vs (NIST 2012)

These guidelines were devised by committees within the respective Standards Office consisting of industry risk practitioners and civil servants with expertise in risk and thus, tend to be practice rather than academic led (BS ISO 31000 2009, FERMA 2003). Further, the guidelines were designed to be an overall framework rather than deal with the operational detail and therefore, although the NIST framework mentions it, what Figure 2.4 does not show is how to categorise the risks identified. This aspect however, forms an important part of identifying and analysing risks because, to properly manage the risk, some form of risk scoring or measurement needs to be defined and assigned to each risk. This can then be used to aid the decision-maker in making an informed decision (Lyon and Popov 2016a).

2.6.1 Measuring Risk

Categories of risk analysis are types of measurement. They may include accessing the likelihood and impact as is the case with the NIST guidelines (NIST 2012). Alternatively, they may equally measure probability, risk level, severity or consequence (Lyon and Popov 2016a). Within IT, other measurements might include criticality, confidentiality, integrity and availability (Heiser 2008). Whichever categories or measures are chosen, to be meaningful, these need to be scored in some way so that decision-makers know how likely a risk is, the potential impact and the consequence. This will then assist in determining how much resource or effort needs to be placed on mitigating or avoiding the risk. Further, scoring may be assigned to risks to bring the information gathered into context for the decision-maker.

Scoring may be done simply, for example, by using a traffic light style scoring mechanism such as high (red), medium (amber) and low (green) or a variation thereof (Hall

2011, Heiser 2008, Lyon and Popov 2016a). Alternatively, risks can be given numeric values. These scores can be calculated as a percentage, e.g. the likelihood of an event occurring is *x* percent. This type of scoring approach is used in project management (Bissonette 2016). Another, way to score risks it to calculate them in terms of financial loss or monetary values (economics). This type of risk scoring is commonly used within organisations to illustrate to management what the cost of a risk occurring will mean for the business. For example, calculating the risk of loss in the case of a financial or investment decision (Virlics 2013); or calculating the cost of brand or reputational damage (Fiordelisi et al. 2013). In relation to data, financial values may be calculated on the cost of a security breach of IT systems (Martin et al. 2014) or a data breach (i.e. a privacy breach), the cost of which can be quite substantial both in terms of reputational damage and financial loss (Ponemon Institute 2015).

For privacy, some suggest risk should be calculated in terms of expected impact (Oetzel and Spiekermann 2014), or harm (Altman et al. 2015). However, attaching monetary values can provide a real focus and help in gaining the attention of senior management as they can see the potential impact as a bottom line cost at a glance. For example, according to IBM, in 2015, the average number of breached records was 21,695 and the estimated cost per capita of a breach of public sector data was $68, while for the financial sector that cost increases to $215 per capita (Ponemon Institute 2015). This shows how monetary measures can be very effective. Further, these figures also highlight the importance of safeguarding privacy by illustrating this is not just a matter of complying with legal regulations, there is also a potential financial cost attached for the organisation, as indeed there should be. Moreover, with GDPR, the fines that a privacy breach can attract have increased to up to 20 million Euro or *”up to 4 percent of total worldwide annual turnover, whichever is higher”* (GDPR, Article 83(4)). This means that the cost of a privacy breach will almost certainly increase. Therefore, because organisations focus on financial profit and loss, money motivates their actions, as can be demonstrated by the propensity to frame risks in terms of what the financial cost of a risk occurring is to the business. Thus, it is suggested, these increased fines will help focus attention on privacy and ensure that organisations will wish to consider privacy risk in much more detail before any data is shared or published in future.

2.6.2 Privacy Risk

Because organisations are accustomed to using a risk-based approach to inform their decision making, it is argued that extending this to also incorporate privacy risks, organisations are more likely to implement and adopt privacy into their decision-making processes. This stance has been adopted in the US, where the recommendation is that any decision-making around privacy should be conducted taking a risk based approach

(Office of Management and Budget 2013), p.195.

A privacy risk, according to ISO, is defined the same as any other risk (BS ISO 31000 2009, BS ISO/IEC 29100 2011). More specifically, a privacy risk is the probability or 'likelihood' and 'consequence' of a loss of, or violation of, an individual's privacy. A privacy violation occurs when privacy is breached or lost, this has been defined by Yale Law Journal as the; *"extent to which we are known to others; the extent to which we are the subject of others' attention; and the extent to which others have physical access to us"*(Gavison 1980). The loss of information privacy occurs when personal information is disclosed; without consent (Borghi et al. 2013); without knowledge (Po-Ching and Pei-Ying 2016, Ohm 2010) or as a result of a data breach (Farrell 2015).

The International Association of Privacy Professionals (IAPP) group privacy into four categories:

1. *Information privacy* relating to collection, storing, handling, processing and sharing of personal information;
2. *Bodily privacy* relating to the physical aspect of individual privacy;
3. *Territorial privacy* relating to the physical environment including personal, home and public;
4. *Communications privacy* relating to any form of correspondence and communications including conversations, letters, email etc. (Swire and Ahmad 2012)

Thus, in relation to information and data, looking at the IAPP categories, the most relevant privacy considerations are information and communication privacy. For privacy perceptions as discussed in Section 2.2, under Westin's functional groupings, the relevant function for data privacy risk will be; *"limited and protected information"* (Westin 1966). Thus, Westin was the first scholar to recognise that privacy is context dependent. Similarly, the other relevant privacy views for data will include the power; *"to control access to, and uses of, places, bodies, and personal information"* (Moore 2015) and affording users the right to self-determine what is communicated about them (Parker 1974) (see Section 2.2).

According to Nissenbaum data privacy is; *"a complex, delicate web of constraints on the flow of personal information that itself brings balance to multiple spheres of ... life"*. The important point is that individuals (users) need to know and trust that information they have supplied, either voluntarily or that organisations have accrued about them is held, handled and processed by organisations in an appropriate manner. Because, although the unwritten privacy rules may still apply, the users are unlikely to apply those rules themselves. Instead the people who process that data will apply the rules on the users' behalf and users need to feel assured that their privacy has been sufficiently safeguarded before data is shared with third parties or even published as open data.

2.7 Frameworks for assessing Privacy Risk

To address this, a search of the guidance and risk frameworks that exist which may provide assistance in determining privacy risks was conducted. This elicited a number of frameworks that provide subject specific guidance around privacy risk. For example, the OECD Privacy Framework considers privacy in relation to trans-border data sharing (OECD 2013), this is based on data protection guidance and aligns with data protection laws in most countries, including the EU and the US. However, EU data protection affords stronger protection than OECD guidelines in that the EU protects data. The OECD protects privacy where the manner or nature of the data processing poses a threat to privacy, whereas EU data protection, affords privacy protection in all processing, whether or not prima facie there is a risk to privacy (Borgesius et al. 2015). Therefore, if practitioners want to ensure those stronger protections of EU regulations are adhered to, EU data protection regulations should prevail. Others that have also considered privacy from a technical perspective e.g. (Barth et al. 2006); from a policy perspective (Bettini and Riboni 2015) and a legal perspective (Borgesius et al. 2015). These studies however, are limited to their particular subject perspective.

The search also found that there are a number of frameworks and guidelines that specifically addresses privacy risk and impact. Details of these is provided in Sections 2.7.1 - 2.7.4.

2.7.1 Privacy Impact Assessment (PIA)

One method for attempting to incorporate privacy into practices is the privacy impact assessment (PIA). In the UK, the Information Commissioner's Office (ICO) recommends that a PIA is conducted for any new project where data handling, storing or processing is involved to assess what privacy implications, if any, may apply. While there are similarities between the DPIA and the PIA, the latter was devised predominantly as a 'tick-box' exercise to act as an *aide memoire* for ensuring privacy considerations are included in project planning (David et al. 2013). Furthermore, in conducting DPIAs, GDPR requires that privacy risk is considered from the perspective of the data subject (see Section 7.2) rather than the organisation. This is perhaps the biggest change that GDPR has introduced as conventionally, when assessing risks, organisations will do so from the perspective of the organisation. Therefore, although some organisations may have conducted privacy impact assessments (PIAs) prior to GDPR coming into force, these are not the same as DPIAs.

Therefore, although conducting a PIA is not a legal requirement, the PIA process was developed and published as a tool to *"help organisations identify the most effective way to comply with their data protection obligations"* (Information Commissioners Office 2014). The steps of a PIA are;

1. Identify the need for a PIA;
2. Describe the information flows;
3. Identify the privacy and related risks;
4. Identify and evaluate the privacy solutions;
5. Sign off and record the PIA outcomes;
6. Integrate the outcomes into the project plan;
7. Consult with internal and external stakeholders as needed throughout the process (Information Commissioners Office 2014).

The PIA appears to suggest a risk-based approach to assessing privacy as it asks practitioners to identify privacy and related risks (item 3). However, like the Privacy by Design (PbD) principles (see Section 2.7.4), the PIA code does not go into too much detail on how to actually apply these principles. It has been argued that although the PIA guidance provides an overview of what to consider in assessing the potential impact of a privacy breach, this is not easy to apply in practice (Oetzel and Spiekermann 2014). Rather, the PIA process is more of a tick-box framework in that it provides a guide to the aspects that require consideration as a reminder for practitioners rather than lead them through a process or direct how best to approach each section.

Moreover, because neither PbD nor DPIAs were obligatory prior to GDPR, these, in terms of CI, are effectively the goal that conducting the risk assessment seeks to achieve, making practitioners think about how privacy can be incorporated as standard before data is shared. Thus, it can be said that PbD is one of the goals or aims of CI but not, as such, an integral part of the framework.

In the US, NIST have approached this from a slightly more detailed manner in their guide to personally identifiable information (PII), which is considered next.

2.7.2 NIST guide to PII

NIST have produced the; *”Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)”* (NIST 2010). This guide also suggests using a risk-based approach to privacy protection (p. 24). Further, it approaches privacy risk from a security perspective, suggesting organisations devise policies and processes covering the following topics:

1. Access controls and rules for data;
2. Data retentions rules and procedures;
3. Incident response including a data breach notification process;

4. Integrate privacy into the system development lifecycle;

5. Data limitation (minimisation) rules and procedures;

6. Breach consequences for failure to abide by privacy rules of behaviour. (NIST 2010)

In addition NIST suggest the following steps are also implemented within the organisation:

1. Training - Devise and run appropriate training sessions to inform staff and ensure they are aware of the processes and rules and their responsibilities in relation to these;

2. Conduct privacy impact assessments (PIA) on information systems to identify and mitigate against risks. The PIA should encompass:

   (a) Information collected - i.e. details of what data is to be collected;

   (b) Reason - i.e. why the data is being collected;

   (c) Use - i.e. how the information is intended to be used;

   (d) Sharing - i.e. how the information will be shared and, if so, with whom will it be shared;

   (e) Security - i.e. what controls and security will be in place to protect the data.

3. Ensure data is anonymised or de-identified (i.e. ensure PII data is obfuscated or removed);

4. Use security controls to protect confidentiality and PII. (NIST 2010)

As can be seen, the NIST incorporates the PIA although this code takes a different approach to how the PIA should be conducted. The US version follows the legal regulations in asking that consideration is given to the information flows, whereas the UK PIA is not particularly data specific. Rather, the UK PIA code asks practitioners to describe data flows and then identify and consider the privacy risks in terms of if there was a data leak or breach. The guidance asks that consideration is given to what the risks would be to; (a) the organisation, (b) the individual and (c) the legal implications/impact but does not clarify how to do this in practice (Oetzel and Spiekermann 2014). Thus, the US approach the PIA process from a system perspective (NIST 2010) whereas the UK PIA is assessing from a project perspective which covers more than just systems. However, both processes are fairly high level and still do not provide much by way of detailed guidance.

Oetzel and Spiekermann (2014) sought to address this when they created a framework for systematically applying a PIA from a system design perspective. This combined the PIA with the NIST risk assessment process to create a more robust risk assessment that incorporates privacy into system design. In doing so Oetzel and Spiekermann (2014) used the data protection principles as privacy targets to be incorporated into the system at design stage, thereby facilitating privacy by design (PbD, see Section 2.7.4). This framework provides a more comprehensive code-based view of how to conduct a PIA in a specific environment, system design. It does however, require *"profound privacy knowledge"* for successful completion (Oetzel and Spiekermann (2014), p. 142). Moreover, the framework does not consider how to assess existing processes or systems, nor does it go into much detail about the context in which the data was collected, handled or the wider context, such as considering the prevailing context versus the changed context if data is released in a new format.

Similarly, Alshammari and Simpson (2018) sought to formalise the PIA process by incorporating system design principles into conducting PIAs by using threat modelling concepts to identify vulnerabilities and threats in order to assess privacy, akin to security threat modelling approaches (Faily et al. 2012). This approach is interesting as it also seeks to incorporate some contextual considerations into the risk assessment process. However, like the Oetzel and Spiekermann (2014) and the system requirements gathering frameworks discussed in discussed in Section 2.7.5, this framework seeks to assess new systems or processes and incorporate PbD at an early stage. Further, to get optimal benefit, users will require detailed knowledge and understanding of threats, vulnerabilities and how to analyse these to establish risk causes and effects. This therefore suggests expert guidance will be needed to use the framework. Further, because this work was published after most of this work was completed, it has not been considered in detail here.

2.7.3 ISO/IEC29100:2011

Another, internationally recognised framework that practitioners can turn to for guidance is the ISO/IEC29100:2011, information technology, security techniques privacy framework. This framework *"provides a high-level framework for the protection of [PII within IT] systems"* (BS ISO/IEC 29100 2011). The UK version of this standard considers privacy in terms of privacy risk in managing privacy risk. ISO/IEC29100:2011 defines the risk management process in four steps. These have been set out in the first column of Table 2.3.

This shows how ISO/IEC29100:2009 can be aligned with Simon's bounded rationality in decision making (see Section 2.5.1) in a similar manner to the comparison to the ISO 31000 risk categories in Table 2.2. The only exception to this is the consultation and communication section which instead aligns with project management principles and the PIA process (Bissonette 2016, Information Commissioners Office 2014).

BS ISO/IEC29100:2011 requires organisations to consider four factors in conducting a privacy risk assessment:

1. Legal and Regulatory - this includes any international or national laws and regula-

BS ISO 29100:2011	**(Simon 1955)**
Context: establish the context by developing an understanding of the organisational and technical environments and what factors, internal and external, may influence privacy risk management	*Context and Unknown*
Risk Assessment: identify, analyse and evaluate privacy risks	*Search*
Risk Treatment identify mitigation strategies and implement privacy controls	*Decide and Act*
Consultation and communication: involve all stakeholders in the process; inform and teach privacy principles and risks to stakeholders.	PIA (Information Commissioners Office 2014)

Table 2.3: ISO29100:2011 vs Simon

tions; court decisions and union agreements;

2. Contractual Factors - this may be agreements with third parties or organisational policies;

3. Business Factors - any specific intended use of the information as well as any industry guidelines or best practice standards;

4. Other Factors - here the standards asks practitioners to consider consent and any internal and/or technical controls or standards that may be in place (BS ISO/IEC 29100 2011).

Further, it goes on to explain the 11 privacy principles:

1. Consent and Choice - processing of personal data should not be done without consent. Further, consent should be informed meaning the data subject should have freely given consent, been informed of and understood his/her rights prior to giving consent and the implications of refusing to give consent;

2. Purpose legitimate and specification - ensuring the data is collected for a legitimate purpose which should be communicated to the data subject;

3. Collection limitation - making sure that the data is processed and handled only in accordance with the legitimate purpose and that any legal constraints are adhered to;

4. Data minimisation - ensuring data is only processed when necessary to meet the legitimate requirements;

5. Use, retention and disclosure limitation - data should not be disclosed or transferred other than for a legitimate purpose nor, should it be retained for longer than necessary;
6. Accuracy and quality - making sure the data is accurate, up-to-date and complete;
7. Openness, transparency and notice - ensuring data handlers (processors and controllers) understand and are aware of legal obligations and the organisational data processing processes, procedures, practices and policies;
8. Individual participation and access - allowing data subjects access to data held about them and to question or challenge the accuracy or completeness of the information;
9. Accountability - ensuring policies, procedures and practices are documented and adhered to. This includes assigning responsibility to designated individuals and establishing a complaints procedure. This needs to be communicated so that data subjects can know how to seek redress if required;
10. Information Security - making sure the organisation and its systems and processes incorporate information security measures on both a strategic and functional level;
11. Privacy Compliance - ensuring processes, procedures and policies adhere to data protection and any other privacy safeguards in place including verifying that any third party who handle the data is also compliant (BS ISO/IEC 29100 2011).

Thus, the ISO/IEC29100:2011 standard provides a comprehensive list of all of the considerations that practitioners must include when considering a privacy risk in terms of information systems. In terms of the data, the framework considers privacy from a predominantly legal perspective in that it follows data protection rules quite closely. To illustrate, a comparison of DPA and ISO29100:2011 principles has been provided in Table 2.4.

This illustrates how this framework attempts to go beyond considering just the impact of a breach. It asks practitioners to consider the wider context and explains the meaning of each term. What is does not do however, is to provide any guidance on how to apply this in reality.

Therefore, while these frameworks provide some very useful guidance, assessing privacy risk according to PIA consists of identifying and managing the likelihood and privacy impact of a breach whereas the ISO standard considers privacy risk as ensuring appropriate policies, processes and controls are in place and followed, and that DPA is adhered to. Like the risk frameworks mentioned in Section 2.6, the formulation of these frameworks was practice led, reflecting practitioners knowledge and insight in security and how to protect data securely. Further, while these incorporate legal protecting of privacy,

DPA	BS ISO 29100:2011
Processed lawfully and fairly	Item 2 - purpose legitimate and specification
Only used for purpose(s) it was lawfully collected for	Item 3 - collection limitation
Used in a way that is relevant and adequate but not excessive	Item 4 - data minimisation
Accurate	Item 6 - accuracy and quality
Kept for no longer than necessary	Item 5 - use, retention and disclosure limitation
Processed and handled in accordance with the data subject's data protection rights	Items 1 (consent and choice), 7 (openness) and 8 (participation and Access)
Kept safe and secure	Item 9 (accountability), 10 (information security) and 11 (privacy compliance)
Not transferred outside the European Economic Area without adequate protection	Item 5 - use, retention and disclosure limitation.

Table 2.4: DPA vs ISO29100:2011

they do not provide systematic practical assistance in how to do this in practice or indeed, how to apply these processes holistically across an organisation.

2.7.4 Privacy by Design

The concept of privacy by design (PbD) provides a more holistic overview of privacy. It was created in the 1990's to *"address the ever-growing and systemic effects of Information and Communication Technologies, and of large-scale networked data systems"* (Cavoukian 2011). In particular, what this framework recognises is that privacy needs to be viewed in a more holistic manner and not just as an exercise by organisations to ensure legal compliance with privacy laws. It consists of 7 steps:

1. *Proactive* rather than Reactive; *Preventative* rather than remedial;

2. *Privacy by default* - i.e. automatic protection of privacy as standard, assuming maximum level of privacy from the start;

3. *Privacy embedded* i.e. privacy should be embedded into the architecture and design of systems, not added at a later stage;

4. *Positive-Sum* rather than Zero-Sum - look for win-win scenarios that affords full functionality while maintaining complete privacy protection;

5. *Full lifecycle protection* - embedding privacy into design of systems and processes;

6. *Transparency and visibility* - involve all stakeholders in open, honest and inclusive communication and processes to ensure everyone adopts the PbD principles;

7. *User-Centric* - respect the privacy of users and data subjects, provide users with privacy notices and user-friendly options (Cavoukian 2011).

What makes the PbD principles stand out however, is that these principles have been recommended as standard good practice within Europe (Danezis et al. 2014) and beyond (Altman et al. 2015). Further, adherence to PbD is now compulsory under GDPR (see Section 2.4.4).

The PbD principles provide a high-level set of principles to follow in ensuring privacy is preserved but do not provide any practical guidance on how to implement this. For example, for specific direction, there are a variety of frameworks that practitioners can turn to for advice on how to assess risk depending on the industry or subject (Bissonette 2016, BS ISO 31000 2009, Lyon and Popov 2016a). One area where PbD has been more developed is within the requirements gathering process for a new system, where there have been a number of frameworks that have sought to implement PbD into requirements and system design. Of these three frameworks are noteworthy.

2.7.5 Privacy Risk Assessment using Goal Modelling

Another method for looking at privacy risk, is through privacy goal modelling. Using goals to derive objectives or requirements has been used in requirements engineering for over two decades (Dardenne et al. 1993). This was extended to anti-goals for security threat modelling (Lamsweerde et al. 2003) and later to define *"desired properties"* that needed protection (Howard and Lipner 2006) as security goals (Faily and Fléchais 2009). More recently, this idea has been extended to include privacy with many of the existing privacy frameworks also review privacy through privacy properties or goals. For example, LINDDUN (*"Linkability, Identifiability, Non-repudiation, Detectability, Disclosure, content Unawareness and policy and consent Non-compliance"*) took two security goal modelling frameworks (STRIDE, (Howard and Lipner 2006) and KAOS (Lamsweerde et al. 2003), and extended this to include privacy (Deng et al. 2011). LINDDUN uses each privacy threat types to map privacy threats to system elements to derive privacy requirements (Wuyts et al. 2014).

Similarly, the IRIS *"Integrating Requirements and Information Security"* (Faily and Fléchais 2009) and CAIRIS *"Computer Aided Integration of Requirements and Information Security"* (Faily and Fléchais 2010) method seek to elicit and visualise security requirements, vulnerabilities and threats using goal modelling, including privacy (Faily 2018). Effectively what privacy goals seek to establish are aims or properties that need protection against a particular action (in security terms 'a threat').

Another example of how security goal modelling has been extended is the *"Privacy Safeguard"* (PRiS) framework, a privacy requirements gathering system (Kavakli et al. 2006) that uses privacy requirements as organisational goals to incorporate privacy into processes and systems (Kalloniatis et al. 2008). These goals are then used to assess and analyse what effect and impact each of the 8 privacy goals (authentication, authorisation, identification, data protection, anonymity, pseudonymity, un-linkability, and unobservability) may have on the system, thereby building privacy into system design (Kalloniatis et al. 2008).

PbD in System Design

The Framework for Privacy-Friendly System Design takes these ideas a bit further by also incorporating different user spheres (user, recipient and joint) into system design. This framework defines privacy according to how *"privacy-friendly"* a system is. System friendliness is derived based on system characteristics which are categorised into privacy stages (0-4) based on *"degrees of Identifiability"* (from identified (0) to, pseudonymous (1-2) and anonymous (3)) (Spiekermann and Cranor 2009). A slightly different approach has been adopted by the European Union Agency for Network and Information Security (ENISA), who have extended on work by Pfitzmann & Hansen and the Independent Centre for Privacy Protection (ICPPS) around privacy terminology and goals (Pfitzmann and Hansen 2010, Hansen 2012, Hansen et al. 2015), to produce guidance around how to align data protection and privacy goals (Danezis et al. 2014). A more in-depth analysis of some of these techniques can be found in Chapter 8, Section 8.5.3 where these concepts have been applied to the data management lifecycle.

These goal modelling frameworks seek to elicit detailed information about how a particular threat or vulnerability might affect the organisation or system and, as such, are very effective for considering privacy in respect of how to ensure privacy is built into the early stages of system design. They are however, less suited for supporting a higher level, holistic overview of privacy decision making. Therefore, while these frameworks adopt an interesting approach to privacy which is well suited to assessing how a particular risk might affect a system or the organisation, they do not provide any form of holistic privacy overview of the risks for what already exists (i.e. existing systems or data) which is the initial area this work is seeking to address.

2.7.6 Privacy-Aware government data release framework

The only framework found that has been designed to consider privacy in the context of open data is the Privacy-Aware Government Data Release Framework (PAGDR), devised by (Altman et al. 2015). This framework has been developed to assist practitioners in designing appropriate privacy policies and adopts a data lifecycle approach to privacy. The framework asks that consideration is given to privacy controls for each stage in the data lifecycle (see Section 2.5).

This framework is intended to ensure that, for each lifecycle stage, the public body considers the following five *"privacy spaces"* and applies appropriate security controls to each 'space', these are:

1. Procedures - this covers devising appropriate data handling processes and vetting procedures;

2. Technical - covering encryption, automation processes and anonymisation practices;

3. Education - covers education of staff and processes for informing data subjects in the event of a breach;

4. Economic - covers economic incentives or impositions such as obtaining insurance and handling potential fines for a breach;

5. Legal - what legal rights will stakeholders, including data subjects, have, how can they obtain redress etc.

This framework provides a more prescriptive, holistic approach but, because it was devised to assist practitioners in ensuring appropriate processes and policies are in place, it does so at a theoretical and fairly high level. This means that the detail of how to implement those policies is not covered as part of this consideration.

Applying these categories to the decision-making process would involve looking at each area within the lifecycle to determine which decisions need to be made at each stage of the data lifecycle. As discussed in Section 2.5, decision-making with regards to open data publishing, requires consideration to be given to whether data currently held and managed can be released in open format. For this decision, what needs to be established is whether the data, in its current format, can be published or, if not, are there any mitigating steps that can be taken, such as suppressing certain fields or anonymisation, that will allow the data to be shared or published. To answer these questions requires a more holistic overview of, not only the attributes within the data but also the surrounding context is needed.

Looking more specifically at data privacy, while most of the frameworks discussed in the sections above consider data privacy, these either consider privacy as one of multiple perspectives (e.g. (Solove 2006, Westin 1966, Mulligan et al. 2016) or they consider how to preserve privacy in a particular domain, e.g. to assess potential impact of a data breach (see Section 2.7.1), or to inform system design (see Section 2.7.5). Others consider privacy as a one aspect of security, treating privacy as a sub-set of setting security parameters around how the data and/or system is safeguarded (see Sections 2.7.2 and 2.7.3) or to determine how a particular threat or vulnerability might affect privacy (see Section 2.7.5).

Thus, although most of the privacy frameworks discussed consider privacy risk, they consider this from one perspective or as one aspect of privacy only. However, the CI framework identified as suitable for adaptation in Section 2.2.1, Contextual Integrity (CI), looks at privacy from a number of different perspectives but places particular emphasis on data and how to assess the privacy risk of data flows in light of the surrounding context and how any change in the flow will impact on the context and therefore, the privacy risks (Nissenbaum 2004 2010)). CI asks that, in considering data privacy, not only must organisations consider the data itself, they must also take into account the context within which the data is processed, who handles the data and how it is transmitted in order to then place appropriate protection around the data. Thus, because CI specifically considers data privacy in detail AND in context, this framework was chosen as the most suitable for adaptation into a privacy-specific decision support framework (See Section 2.2.1). To this end, CI is explained and discussed in more detail next.

2.7.7 Contextual Integrity

Nissenbaum introduces the Contextual Integrity (CI) framework and considers the importance of putting privacy into context. This involves considering all the nuances of privacy including:

- Access - how data is accessed (*"Access Definitions"*);
- Control - how data is controlled by those public body who collate the information and the surrounding legal and policy frameworks (*"Source of Prescriptive Power"*);
- Context - consideration of what privacy is within a given context (the *"normative and descriptive conceptions"*).

Privacy in relation to data and information is described as; *"a right to appropriate flow of information"*. Thus, to apply this to information and navigate the information flows, CI asks practitioners to consider the information flow, looking at it from three perspectives; Actors; Attributes and Transmission Principles.

Actors, are the; *"subjects, senders and receivers"* of the data. This refers to the people that are the subjects of the data itself (*"the information subjects"*) and those who handle the data (the *"data senders"* and the *"data receivers"*). The attributes refer to the individual elements that make up the data, this can be described as the rows and columns within a database, each containing pockets of information (attributes). Finally, the transmission principles refer to how the data is conveyed and shared, i.e. the data flow between the actors.

Different actors will view the information from different perspectives. What is considered sensitive by an individual may not be considered so by the data controller within the public body. For example, the type of ailment a person has may not be considered sensitive by the data controller when collating information about types of diseases in a particular area. However, for the individual who perhaps is one of only a few people with that ailment who resides in that area, this information can lead to them being re-identified or even refused health insurance. Thus, the individual may consider such information very sensitive and, by the same logic, similar arguments can then be derived around attributes and transmission principles.

For the data itself, this is defined as attributes. Attributes are the individual information elements that make up the data, this can be described as a table, containing rows and columns within a database. Each column will contain pockets of a specific type of information (attributes, for a more detailed description see Section 2.4.3). Finally, the transmission principles refer to how the data is conveyed and shared, i.e. the data flow between the actors.

Further, to cover the context, each actor also needs to be considered in context. This means that each actor will need to be considered in relation to the role the play. Thus, the data controller will need to be evaluated in relation to their job role, e.g. a data analyst, or social role such as student. CI then considers the activities those roles may carry out, e.g. analysing data, being taught, what duties, prerogatives or obligations are associated with each role, what defines the norms (behaviours), their trustworthiness etc.. Norms may be implicit or explicit, so they may be prescribed norms or established customary etiquette. Finally, the values, purposes, goals and ends of the particular surrounding setting or situation are considered.

Existing Applications of the CI Framework

There are examples of how the CI framework can be applied in the literature. For example, (Barth et al. 2006) sought to incorporate context into privacy decision making by devising a framework for controlling data based on CI, that will compute what information users can access and/or share with others. The framework seeks to achieve this based on system settings and access controls. The framework focuses on the flow of information between roles, taking into consideration the attributes within the data relating to a particular individual (data subject), and the role played by the actor who handles the data at that time. Values and norms are incorporated into what values and/or norms a particular role will hold, rather than the values and norms of the actor who performs that role.

To do so, Barth et al. (2006) incorporates the values and norms of CI into the roles of the individuals (such as doctor, lecturer and any derivative or associated roles) by pre-setting or computing into the framework the values (aims, ends and goals) which that particular role is expected to hold in relation to the information they handle. This association is then used as one of the considerations that sets appropriate access controls for that role. The other controls set by the framework seek to incorporate controls that will allow any policy and legal constraints that could influence whether role A can share particular data with role B to also be taken into account. Thus, arguably, while Barth consider values and norms at an organisational (policy) and governmental (legal constraints) level, they do not sufficiently consider the wider context or the multiple values and norms that each actor will hold. To illustrate, an actor is a human being who will, in relation to handling the data be; the subject of (*data subject*); the sender; or the receiver of information. The actor will perform one or more roles and bear different responsibilities for each role. Thus, the actor will perform an action with the data (subject, sender or receive) and as part of that performance, be attributed to a role which holds certain values and norms. However, the actor will also hold values and norms as a person and this aspect should not be ignored. For example, an actor may be the sender of information in their work role as a doctor. The same actor may also be a patient, friend and/or colleague to the data subject at the same time and this may alter the context of the data flow depending on the setting. These nuances also need to be allowed for and considered as part of any application of CI.

It is argued, that in trying to box CI into a computational environment, the Barth framework considers CI in-depth at too early a stage. Before controls can be computed and set, a holistic privacy assessment needs to be conducted. The Barth framework considers the data relating to one individual's information, rather than for a group of individuals. Moreover, it considers this information based on the information flows (transmission principles) between the roles that handle the information, setting appropriate controls derived from access rights and any policy or legal restrictions that may apply, thereby failing to sufficiently consider the contextual nuances between actors and their multiple roles. Further, while Barth's framework enables roles to collect and share an individual's data with known third parties (roles), it does not allow for data being published or shared openly with unknown third parties (Barth et al. 2006).

Other applications of CI consider applying CI through tags attached in the message headers (Krupa and Vercouter 2012) or have applied CI to a particular question or problem and, in doing so these studies discuss whether privacy is preserved in that scenario. For example, the CI framework has been applied to cloud storage and social networking sites to determine whether these sites provide sufficient privacy protection (Sar and Al-Saggaf 2013, Grodzinsky and Tavani 2011). What these studies all have in common is that they have been applied to scenarios where all the elements of the framework are present, i.e. the roles, context and transmission principles can be specifically defined. Moreover, these have been considered either without input from the data controller/processor and/or sending party, or, they have been applied on a to a theoretical problem.

Another study (Conley et al. 2012), sought to apply CI to the open data publishing domain, the area of focus for the first case study of this work (see Section 1.1.1). This study reviewed the merits of making court records available online by comparing a manual search of the records with an online search. The study did not seek to systematically apply CI, rather it discussed the different aspects of CI in the context of making the data available in open format. The study concluded that the data should be either anonymised or restricted access should be applied in light of the privacy implications found. However, whilst this study considered the domain of open data publishing, it did so from a case study perspective considering the merits of making data publicly available rather than guide practitioners through the process of applying CI to the decision-making process, which is the focus of this work, starting with making privacy decision for open data (see Chapters 5 and 6).

Thus, on a theoretical level, it appears that the CI methodology enables an organisation to strategically consider the privacy implications of a dataset both from an individual attribute level and from a contextual perspective. It therefore offers an excellent basis from which a more detailed model can be created based on the principles of CI.

To this end, it is contended that the contextual integrity (CI) framework can be adapted to provide the solution. CI combines the holistic overview of PbD with context and, as such, lends itself well to be adapted for practical use in organisational decision making around privacy risks. Further, by incorporating context into the privacy consideration, the CI framework ensures any contextual interference or influence is also accounted for to make the assessment reflective of life and how context can influence our behaviours and actions, and thus, any resulting privacy assessment will be more through. For those reasons, the initial choice that made CI appear ideal and suitable for adaptation for this work have been verified and thus, CI has been chosen as the framework of choice to be adapted for the privacy-specific assessment framework that will be created in this thesis.

2.8 Literature Review Conclusion

This review has looked at how privacy is defined (Section 2.2), determining that privacy is viewed differently depending on the perspective lens of the reviewer or assessor. The review then went on to consider data privacy (Section 2.2.1), and established that, although many of the academics consider data privacy as one aspect of privacy, one framework that centres more particularly on data privacy is the Contextual Integrity (CI) framework. This framework was therefore chosen as suitable for adaptation to answer the research questions (see Chapter 1, Section 1.2).

In view that the open data domain was chosen as the starting point for this work (see Section1.1.1), the review went on to provide a brief overview of the path to open government and open data (Section 2.3), to provide a backdrop for the focus for the first study which will scope how public bodies currently consider privacy risk and open data publishing (see Chapter 4).

The literature review went on to discuss how privacy can be protected in light of conflicting obligations in sharing or publishing data, whilst seeking to maintain confidentiality and privacy of the data subjects whose data may be held or processed by an organisation (Sections 2.4 and 2.5). This looked briefly at practical and technical protection, as well as a more detailed discussion of both the past and current legal regulations applicable in the UK over the lifespan of this thesis, the Data Protection Act 1998 and GDPR and how these can be applied to CI.

Resulting from this, data privacy in respect of organisations who own datasets containing personal data (user data), can be defined as: *maintaining the confidentiality of user data held, managed and processed by the organisation*. Further, while legal, practical and technical protection options exist for preserving privacy, it is by no means standard practice whether and how this is applied.

This led to a review of decision theory and how organisations make decisions which concluded that a risk based approach is often used in decision making. Further, a risk based approach is an effective way to assess privacy and help inform privacy decision-making within an organisation, as can be evidenced by the fact that existing privacy decision-making frameworks adopt a risk based approach to determine the likely privacy implications of a particular scenario (Section 2.6). Therefore, using a risk based approach for creating the privacy-specific assessment framework for this work seems highly appropriate, as organisations will be familiar with such methods. The review showed that there are a number of risk assessment frameworks that practitioners can use to assess privacy risk, including some privacy specific frameworks (Section 2.7). However, most of these are domain specific (e.g. privacy goals or system design, see Section 2.7.5), or security frameworks adapted to better incorporate privacy risk (e.g. Sections 2.7.2 and 2.7.3).

One framework that provides a more holistic overview is the Privacy by Design (PbD) (see Section 2.7.4), however, while this framework covers all aspects of privacy the organisation will need to consider, it provides little specific guidance in how to implement this in practice or how to account for any contextual nuances or influences and how these might affect or alter the flow of the data or indeed, the people who handle the data on behalf of the organisations. As discussed in Section 2.7.7, the Contextual Integrity framework, combines the holistic overview of data privacy with both context and privacy risk. Thus, Contextual Integrity (see Sections 2.2.1 and 2.7.7) was chosen as the framework of choice to be adapted to answer RQ1. The reason for choosing this framework was that CI appears to cover all the aspects discussed. What is also evident from this review is that contextual integrity perhaps can offer a more in-depth analysis option that would consider not only the legal and process oriented view, but also put privacy into context and this is the aspect that this work will seek to explore in the first instance.

Chapter 3

Methodology

3.1 Introduction

From the literature review it was established that businesses make decisions based on bounded rationality when looking for suitable solutions to a problem (Simon 1955), which, in practical terms, often involves looking at the risks associated with each possible solution. Thus, risk forms an intrinsic part of decision-making in business (Galanc et al. 2016). For privacy however, while guidance exists, it is predominately practice led and does not contain much practical guidance on how to assess privacy risk in practice to enable informed privacy decisions to be made in regard to the data being shared or published.

This work seeks to seeks to address this gap by looking at how effective a risk based decision-making approach can work in assessing privacy risk and create a framework that will clarify and assist practitioners in making informed decisions. In seeking this clarification, this study aims to leverage on existing frameworks to validate or otherwise how effective these are in practice. Based on the results of this, the aim of these studies will be to create a privacy assessment framework, trialled in practice, that practitioners can use to support privacy decision-making.

Starting with an explanation of the different approaches to research design and choice of methodology, this chapter will outline the research design and methodology for this study. Choosing a research design and methodology is an important aspect of any research project as this will guide and direct the whole project. The remainder of this chapter will seek to explain these terms, starting with paradigms in Section 3.2, where five leading research paradigms are defined: Positivism (Section 3.2.1); Post-positivism (Section 3.2.2); Interpretivism (Section 3.2.3); Critical inquiry (Section 3.2.4) and Pragmatism (Section 3.2.5). This is followed by and explanation of the three different categories of research design available to researchers in Section 3.3. Next, in Section 3.4, the areas which should be considered as part of choosing suitable research design and methodology outlined. This section includes an explanation of common methods used in social research: Biography (Section 3.4.1); Phenomenological Studies (Section 3.4.2); Grounded Theory (Section 3.4.3); Ethnography (Section 3.4.4); and Case Study (Section 3.4.5). In addition a number of techniques are described that have been used as part of this work: Semi-structured interviews (Section 3.4.6); Contextual interviews (Section 3.4.7); Think-aloud (Section 3.4.8); Story Telling (Section 3.4.9); Scoping Study (Section 3.4.10; and Prototyping (Section 3.4.11). The chapter concludes a discussion and consideration of each area discussed in turn before selecting the chosen methodology, a case study, as the overall methodology for this study in Section 3.5 and explaining the reasons for this choice. Section 3.5.1 provides the protocol that will underpin this case study.

3.2 Paradigms

A methodology or research design should be based on a paradigm or worldview that lends itself to the topic or phenomenon being studied. This can then be supported through the selection of suitable methods, procedures or strategies that will support and structure the data gathering, analysis and interpretation of the research conducted (Creswell 1998).

A paradigm is *"a basic set of beliefs that guide the action"* Guba (1990), p. 17), and this will likely be shaped by the study discipline area, the experiences of the researcher(s) and their peers (Creswell 1998). This means that the approach chosen for a study will be derived from a philosophical paradigm or worldview brought to the study by the researcher(s) conducting the work as this will influence the research and therefore, must be identified (Creswell 1998).

There are many paradigms that guide our actions ranging from religious paradigms that guide our spiritual life through to those that guide a structured inquiry such as a research study (Guba 1990). Each paradigm will have a view on the ontology, epistemology and methodology and therefore, according to (Guba 1990), the choice of paradigm can be guided by answering three questions. Table 3.1 sets out definitions for each term as defined by (Schwandt 2001) (column one) and the Guba's questions in column two.

Definition adapted from (Schwandt 2001)	**Question adapted from (Guba 1990)**
Ontology Concerns comprehending the *"things that constitute the world"* (p. 190)	*"What is the nature of knowledge?"* (p. 18)
Epistemology "The study of the nature of knowledge and justification" (p. 87)	What is the essence of the relationship between the researcher and the known?
Methodology "the theory of how inquiry should proceed" (p. 193)	How should the researcher approach discovering the knowledge?

Table 3.1: Paradigm views and questions that guide the researcher's choice of paradigm

Thus, each paradigm has a set of different viewpoints and by answering Guba's questions the research can select which paradigm is best suited to the study being conducted. The paradigms consider here are: positivism, postpositivism, interpretivism, critical inquiry and pragmatism, each will be defined in turn.

3.2.1 Positivism

Positivism concerns explaining or discovering the reality of how the object of research really is and how it *”really works”* (Guba 1990). It reflects the opinion that where a problem exists, there is a solution (Arghode 2012). Thus, the researcher should adopt an objective approach and so be detached from the object of research. Positivists seek to capture ’reality’, as closely as possible, without altering it with the aim of scientifically verifying, explaining or proving knowledge or a theory, *”leading to control and predictability”* (Blaxter et al. 2010). The ontology, epistemology and methodology views of positivism can be found in Table 3.2.

View	**Definition**
Ontology	Based on realism, i.e. that a reality that is driven by immutable natural mechanisms and laws exists in the world (Guba 1990)
Epistemology	There is one *”single truth”* (Arghode 2012)
Methodology	Experimental inquiry using hypotheses and/or questions (Guba 1990). i.e. suited to quantitative methods that can be measured without necessarily requiring subjective enquiry e.g. using controlled experiments or statistics

Table 3.2: Positivism paradigm beliefs

3.2.2 Post-positivism

View	**Definition**
Ontology	Critical realism, i.e. while there is a reality that is driven by immutable laws and mechanisms, this can *”never be fully comprehended”* ((Guba 1990), p. 23)
Epistemology	Approximated objective inquiry, i.e. objectivity remains the ideal however, this can only ever be approximated (Guba 1990)
Methodology	Utilising amended experimental inquiry to conducting research in natural settings (Guba 1990), i.e. suited to quantitative methods

Table 3.3: Post-positivism paradigm beliefs

Post-positivism represents an amended view of positivism in that post-positivists accept that subjectivity may affect observations. Post-positivists thought argues that complete objectivity is not possible and thus, while post-positivism has the same basic beliefs and aims of positivism, post-positivists accept that the values and background knowledge of the researcher(s) can influence the research. Researchers therefore must acknowledge these shortcomings by checking and validating their findings. To expose ’reality’, post-positivists believe that critical examination of a wide selection of subjects and or evidence will be necessary as they seek to study and explain probable causes which influence outcomes through careful measurement and observation. Thus, like positivism, this paradigm lends itself better to quantitative enquiry (Creswell 1998). Table 3.3 outlines the ontology, epistemology and methodology views of post-positivism.

3.2.3 Interpretivism

Interpretivism concerns understanding a phenomenon as opposed to explaining it (Blaxter et al. 2010). Interprevism may be referred to as ’Social Constructivism’, advocates the view that people want to understand the world and develop meaning from their interactions and experiences (Creswell 2009).

Using subjective enquiry to examine the object of research through observing actions or eliciting views and opinions of research subjects, these paradigms seek to obtain findings through interpreting and analysing these actions, views or opinions rather than through measurements. Social Constructivists (also known as Interprevists) believe people form views and opinions about the world around them and, from those, meaning can be derived. Interpretivism and social constructivism therefore, can be described as a method of enquiry to qualitative research (Creswell 1998). Table 3.4 describes the ontology, epistemology and methodology views of interpretivism.

View	**Definition**
Ontology	Seeking explanation *”by understanding the reality”* (Arghode 2012)
Epistemology	Sensitive observer intersubjectivity in understanding the *”multiple truths”* (Arghode 2012)
Methodology	Observation or questioning of participants. i.e. suited to qualitative methods such as case studies and semi-structured or contextual interviews

Table 3.4: Interprevism paradigm beliefs

3.2.4 Critical Inquiry

View	**Definition**
Ontology	Critical realism (as in postpostivism)
Epistemology	Subjective inquiry where *”values mediate inquiry”* (Guba 1990)
Methodology	The process of argument - using comprehension and evaluation of the texts to construct an argument (Boylan 2009), i.e. suited to qualitative methods such as narrative inquiry or critical reflection

Table 3.5: Critical Inquiry paradigm beliefs

Critical inquiry (or critical theory) concerns values and ideals, it seeks to challenge existing conventions and therefore, can be associated with causes *"that seeks to bring about change"* (Blaxter et al. 2010), such as anti-racist, homosexuality studies or neo-Marxism. Critical theorists seek to construct arguments that can *"properly persuade others"* (Boylan 2009). The ontology, epistemology and methodology views of critical inquiry can be found in Table 3.5.

3.2.5 Pragmatism

The major principles of pragmatism are that truth is never absolute, it is relative and the human organism consists of mind, body and soul, none of which can be exist without the others (McDaniels 1983). Pragmatism concerns practically applying ideas through acting on them and testing them in human experiences (Gutek 2009). Pragmatists focus on application rather than method and without the constraint of having to commit themselves to one particular method, allowing both quantitative and qualitative assumptions to form part of the research (see Table 3.6).

View	**Definition**
Ontology	Rooted in realism, the notion of a complex, multilateral reality, thus *"we construct our worldview based on our perception of it"* (Houghton et al. 2012)
Epistemology	Psychosocial - using human experience and perception to elicit insights
Methodology	Application rather than method - therefore, supports mixed method research as it provides researchers with the freedom to apply multiple methods, paradigms and assumptions as well as different types of data collection, interpretation and analysis (Creswell 1998)

Table 3.6: Pragmatism paradigm beliefs

3.3 Research Design

Research design can be grouped into three categories: quantitative, qualitative and mixed method. These can be described as:

Quantitative: focuses on numbers and structured information. Useful for explaining phenomena that can be quantified and measured in some way, e.g. numerical, statistical or mathematical. *"Quantitative research tends to involve relatively large-scale and representative sets of data, and is often... presented or perceived as being about the gathering of 'facts'"* (Blaxter et al. (2010), p. 65). This form of study lends itself to closed-ended questions.

Qualitative: focuses on words or less structured information, sometimes referred to as 'rich data' because it; *"describes the notion that qualitative data ... reveal the complexities and richness of what is being studied"* (Given 2008). Useful in studies where understanding of observations, behaviours or 'rich data' are collected and analysed. Qualitative enquiry involves collecting and analysing information in *"as many forms, chiefly non-numeric, as possible"* (Blaxter et al. (2010), p. 65). This type of study therefore, lends itself to open-ended questions (Creswell 1998). According to de Ruyter and Scholl (1998), qualitative research affords the use of small samples, from which in-depth insight into participants views can be gained. It provides insight rather than measurement and thus, is a flexible exploratory approach from which results obtained are realistic, concrete and ideal for idea generation (de Ruyter and Scholl 1998).

Mixed methods Represents the middle ground where both quantitative and qualitative methods are incorporated into the study design (methodology). It requires using philosophical assumptions and mixing of both quantitative and qualitative approaches in parallel as part of the research. That said, it is possible to utilise an overall methodology that is quantitative or qualitative in nature and still use both types of enquiry as individual of research methods within the main body of the research.

3.4 Choosing of Methodology

A Methodology is the overall plan or strategy for conducting the research while methods are the individual research tools or lines of enquiry applied as part of the study. That said, one or multiple methods will have to be chosen as the overall methodology that frames the study. That does not mean that only one method may be used as part of a study however. Depending on the method(s) chosen, multiple methods may be utilised as part of the study beneath the main methodology. Similarly, it is possible to use mixed methods within a study and combining quantitative or qualitative research techniques within one or more of the underlying projects that form part of the overall study. For example, a study conducted as part of a research project may collect some data that can be analysed both numerically and qualitatively such as a questionnaire. Here, it would be feasible that some information collected, e.g. participant demographics, may be analysed statistically (quantitatively) while, other parts of the same questionnaire may be analysed based on the words used (qualitatively).

Selecting a methodology requires thought be given to a number of elements and, to this end, (Blaxter et al. 2006) poses eight areas that should be considered: the purpose of the research; the role of the researcher; the nature of the knowledge; how will the research be assessed; ethics; writing style; methodology and method and who the intended audience are. Each will be considered in turn.

Research Purpose The phenomenon being researched here is a theoretical problem of how to facilitate practitioners (humans) being able to make decisions around and incorporate an idea or ideal (privacy) into their business processes, in order that they can make informed decisions about privacy;

Role The role of the research in this work will be *”as an instrument of data collection”* (Creswell 1998) in the form of an observer, information gatherer and facilitator. Gathering information and documentation from existing research and publications, observing the characteristics of existing methods and how these are/may be applied, gathering views of participant to gain understanding and from that, establish meaning that can be translated into new methods;

Knowledge Ultimately, all knowledge is based on human experience and how they perceive the world, the research and knowledge of the past has formed the basis for the facts that are now known. Thus, this work will follow the interprevist ontology by first ”understanding the reality” (Arghode 2012), in order to produce new knowledge in the form of a framework that will aid practitioners in better decision making around privacy and how best to preserve privacy;

Assessment Criteria The research will be subjective as it will consist predominantly of rich data. However, the analysis of the data will be conducted from an objective perspective, i.e. protocols will be written and followed prior to conducting any studies and findings will be validated to avoid bias, enable replicability and ensure procedural processes have been followed. The data collected will be analysed and validity will be sought through verification, evaluation and triangulation of research and findings. Rigorous data collection and analysis will be employed consisting of multiple forms of data. This will involve conducting interviews, observing existing methods and/or behaviours and studying whether there are any existing tools and/or support mechanisms that can be applied to solve this problem. Thus, the work may be described as; *”an inquiry process of understanding ... that explore[s] a social or human problem.... in a natural setting”* ((Creswell 1998), p. 15) and, for that reason, the research will lend itself best to a qualitative method of inquiry. The contribution to existing knowledge will be *a privacy decision making framework that is useful, insightful, informative and accessible to practitioners, so as to encourage and facilitate better privacy decision making processes as standard within organisations*;

Ethics Ethical approval will be sought from the University Ethics Committee prior to approaching any study participants. Details of applications, participant information sheets, questions and any questionnaires and approvals will be included in the study. The research will be conducted with critical subjectivity involving confidentiality, reciprocity and mutual respect between the researcher and participants and will be assessed through evaluation and triangulation. Any bias will be explained and clarified as part of the individual sub-studies carried out as part of this study;

Style The writing style for this work will be academic. Academic writing style uses formal and impersonal styling. Findings will be expressed clearly and concisely as possible, avoid being emotive language and seek to be objective rather than subjective;

Methods There are many methods or research techniques that can be employed as part of research design. Creswell (2009) recommends that for qualitative research, one of five traditional methods is best suited to be used as the overall methodology: a biography, a phenomenological study, grounded theory, ethnography and case studies (p. 176). Thus, these five methods have been explained briefly in Sections 3.4.1 - 3.4.5;

Audience The aim of the work will be to *"transform research findings ... into academic knowledge"* ((Hyland 2004), p. 6) and thus, the audience will be academic peers and anyone interested in privacy risk and decision making around privacy.

3.4.1 Biography

Creswell (1998) describes a biography as the study of one individual and their experiences as found in archival documents, or told to, the researcher. The output may be an autobiography, an individual biography, a life history or an interpretive biography as the researcher creates the person they are writing about *"just as they create themselves when they engage in storytelling practices"* (p. 48). The method of conducting a biography will consist of using an interpretive approach, collecting and digesting extensive stories and information about the subject to develop a detailed knowledge about the context, historical timeline and position of the subject within society (p.51).

3.4.2 Phenomenological Study

Rooted in philosophical ideas, phenomenological study methods have been used in human and social sciences, e.g. health sciences, sociology and psychology. Phenomenologists study social actions and how people perceive the world. Central to this is the idea of epoche, meaning the researcher needs to seek understanding from the views of participants and remove any personal subjectivity or points of view from the study. A phenomenological study depicts meaning from the *"lived experiences"* of multiple individuals about a phenomenon or concept seeking the essence or *"underlying meaning of the experience"*. Phenomenological methodology involves analysing data through reduction, analysing themes or particular statements, and searching for possible intimations (Creswell 1998).

3.4.3 Grounded Theory

As the name suggests, grounded theory (GT) is based on building a theory from the ground up on the phenomenon or concept being studied, in order to generate a theory or hypotheses, i.e. GT is an inductive approach to research, it is concerned with generating theory rather than testing theory (deductive approach). The method involves using rich data, such as interviews, and processing these through a series of coding exercises (open, axial and selective coding), and grouping information into categories. Each coding phase will narrow down the categories until *"saturation"* is reached and one or a few categories remain (Corbin and Strauss 2008). This will then identify *"a story line"* which can be used to generate theory or hypotheses (Creswell 1998).

3.4.4 Ethnography

A study or investigation of a social phenomenon or problem where the researcher immerses themselves into the lives, group, situation or culture under study. Researching the phenomenon is carried out through materials, interviews, participation, language, observation and interactions and the study of the meaning of these behaviours, interactions etc. The aim is to discover pervasive patterns or themes in the daily lives of the study subject(s) that provide useful insights. This method is complex and requires the researcher to invest considerable time in the study, predominantly in the field and have a grounding in anthropology (Creswell 1998).

3.4.5 Case Study

Case studies concern exploring a "bounded system" (a.k.a. a case) over time using in-depth, detailed data collection, involving rich data from multiple sources (Creswell 1998). A case study can be used to; test theory, provide description and/or generate theory (Eisenhardt 1989). It can be defined as:

> *"an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident"* (Yin 2013).

The object of the case study, known as the *"unit of analysis"* (Yin 2013), can be an event, a programme, individual(s) or it may be an activity or issue and the case study can incorporate different sources of information which may include observation, documentation, physical artifacts, interviews and audio-visual materials. It is also possible to use both mixed methods and multiple methods within a case study (Creswell 1998, Yin 2013).

In addition to these five methods, there are a number of research techniques that can be used to compliment or support the main method chosen as the overarching methodology. Some of these will be used to support the selected methodology for this work (see Section 3.5) and therefore, these are also described briefly in Sections 3.4.6 to 3.4.11.

3.4.6 Semi-structured interviews

Semi-structured interviews are interviews where the researcher engages in formal interviews by asking a pre-determined list of open-ended questions to participants, following an interview guide, i.e. a list of topics and questions that need covering during the interview (Cohen and Crabtree 2006). The intention being this keeps the participants on subject while still allowing participants considerable leeway in their responses (Gagnon 2010). However, while the questions are pre-determined, in a semi-structured interview, the interviewer can follow trajectories that come up as part of the conversation that may stray from the guide where this is considered appropriate by the interviewer.

Conducting semi-structured interviews involves developing an interview guide that should include: identifying the need for using semi-structured interviews as the chosen technique and using previous knowledge and/or a literature review to guide the questions. Further, the questions should be tested and presented in the final semi-structured interview guide (Kallio et al. 2016).

3.4.7 Contextual interviews

Contextual interviews involve one-to-one interviews with participants in-situ, e.g. in their workplace, observing them while they carry on with work. The contextual interview technique requires that suitable participants are targeted who carry out the activity being studied on a regular basis. The principles involved in contextual inquiry are;

1. Context (study users in their work);

2. Partnership (talk to users about their work asking about the unarticulated aspects of their work);

3. Interpretation (develop a shared understanding with the user about which aspects of the work matters); and

4. Focus (direct the questions from a clear understanding of your own purpose in being there) (Beyer and Holtzblatt 1998).

The interview technique centres on observation rather than conducting the interviews, the interview element is intended only as a technique for further explanation or elaboration to clarify what is being observed (Holtzblatt et al. 2005). This technique is most commonly used in user centred design.

3.4.8 Think-aloud

Think-aloud as a research technique has its roots in cognitive psychology (Ericsson and Simon 1980). The technique involves a study participant verbalising their thoughts as they carry out a task, to allow the researcher to record what the participant is experiencing through their own account (Ericsson and Simon 1980). This allows information about how the participants solves problems to be gathered (Fonteyn et at 1993).

There are three types of verbalisation probing according to Ericsson and Simon; (a) Talk-aloud or think-aloud - this involves the participant verbalising their actions or thoughts as they carry out a task; (b) Concurrent probing - here participants are probed for specific information as they perform the task (c) Retrospective probing - this involves asking the participant to verbalise what happened after performing the task (Ericsson and Simon 1980).

Research has found that where think-aloud data is combined with *"retrospective thinking"* obtained in follow-up interviews, an accurate and detailed picture of the participants reasoning in problem solving can be obtained (Fonteyn et al. 1993).

3.4.9 Story Telling

Storytelling as a research technique is used to understand users. It involves collecting stories or narratives to generate output. This may involve collecting narrative analysis whereby the researcher will recount a narrated event to produce a story, or *"paradigmatic"* analysis of narratives or stories to conceptualise occurrences, tasks, ideas or events and produce categories of typical occurrences (Bruner 1986). This technique has been used in Psychology (Bruner 1986), software design (Desilets 2008) and IT and business research (Rooney et al. 2016) to name but a few.

3.4.10 Scoping Study

Scoping studies are a form of literature review which, rather than consider an area in detail using recognised academic sources, will utilise other techniques and sources to investigate an area of interest or question or in order to gain an overview of *"the key concepts underpinning a research area and the main sources and types of evidence available"*(Arksey and O Malley 2005), which can then be used to eliminate or inform areas for further investigation or study.

3.4.11 Prototyping

Prototyping has been used as a research tool since the late 1980's. It involves the creation of a system or tool that can, for example, be used to; understand a particular problem; summarise knowledge; test theory; or better understand human decision making. It can also be used as part of case study research to help provide a practical tool that can act as *"proof of concept"* for an idea or system O'Leary (1988). In computing prototyping can be done as a computer- or paper based proof of concept for the software or tool being developed, both types of prototyping have proved equally effective in practice (Walker et al. 2002, Sefelin et al. 2003).

3.5 Selected Methodology

For this work, the methodology selected is a case study. The reasons this has been chosen as a methodology are:

- The case study supports the interpretivism paradigm in that it seeks to gain knowledge through understanding the reality and the *"multiple truths"* (Arghode 2012);
- It supports qualitative enquiry, collecting and analysing 'rich data' (Creswell 2009, Blaxter et al. 2006, Yin 2013);
- A case study can be layered to allow multiple levels and methods of analysis within the same study (Eisenhardt 1989), using multiple research techniques can; "provide a richer picture of the events" (Sawyer 2001) than any alternate single method could;
- It is one of the five methods recommended as particularly suited for selection as overall methodology (Creswell 2009).

This case study will be follow the methodology of (Yin 2013), which involves ensuring five areas are included in the methodology (Yin et al. 1985):

1. Problem definition: defining the case (Section 3.5.1);
2. Protocol (research design): create protocol defining the propositions, research questions and research strategy (Section 3.5.1);
3. Nature of evidence: data collection (Section 3.5.2);
4. Analysis and interpretation: define analysis approach (Section 3.5.3);
5. Manner of presentation: how findings will be presented (Section 3.5.4).

3.5.1 Case Study Protocol - Problem definition

The aim of this research is to conduct an explorative case study of how organisations make decisions around privacy when sharing and publishing information and, in particular, when data is published as open data. An explorative case study has been found effective for organisational practice scenarios where multiple sources of evidence are utilised (Yin et al. 1985).

In order to understand the privacy decision-making process, this work will explore how decisions are currently made rather than explain how they should be made. By getting an understanding of existing methods and what guidance, if any, is used, it will be possible to establish what level of support is available and how effective this is perceived to be in practice.

The literature review concluded that data privacy refers to maintaining the confidentiality of user data while under the control of the organisation (see Chapter 2, Section 2.8). Where data is published or shared, this protection will need to be extended to any third-parties who handle the data. The review outlined the existing methods and thinking surrounding organisational decision making, privacy, risk and, in particular privacy risk and concluded that a risk based approach to the privacy decision-making process can be effective in incorporating privacy into the decision making process prior to open data publication in practice (see Section 2.8).

The review showed that there are a number of existing frameworks available for supporting decision-making around privacy (see Section 2.7). Most however, have not been tested in practice and none were found to have been trialled in respect of privacy decision making around open data release. Therefore, the second goal of this study, will be to trial contextual integrity (CI, see Section 2.7.7), an existing framework, in practice to ascertain whether this framework can provide effective privacy decision support for open data publishing.

From this it will be possible to establish the effectiveness of existing frameworks and create a revised, amended or new framework to further facilitate and support decision making around privacy. As part of this, a review of contemporary legislation will be necessary to ensure these provisions are incorporated into any framework or guidance documentation so that sufficient heed is afforded to data privacy in decision making in future frameworks/guidance.

Research Strategy

The choice of subject in this work has been influenced by the prior knowledge and experience of the lead researcher and therefore, the worldview or paradigm that is most congruent with this work is interpretivism. Interpretivism recognises that the researcher's prior knowledge and experience influences decisions and takes account of this by allowing the researcher to focus on the interactions or processes within a specific context (Creswell 2009), in this case privacy, within the context of how practitioners make privacy decisions around data sharing or publication.

The aim of the research is to establish how privacy practitioners can best manage the tension between allowing access to data in a manner that will retain data utility and preserve user confidentiality, and the privacy concerns that open data release will bring. For this it is expected that the data collected will come from a variety of documents, these may include theory, interview transcripts, recordings and questionnaires collected in the course of this research, i.e. predominantly rich data. This aligns with the interpretivism paradigm of seeking explanation though comprehending the reality and understanding the *”multiple truths”* of a phenomenon (Arghode 2012), in this case privacy decision making. This paradigm was chosen because it reflects the view of the researcher that all knowledge is historically and socially informed in that it is based on interpretations and understandings of people over time. Further, perceptive reality can be elicited from dialogue and investigation and this can uncover ’truth’ and from this, findings can be generated. Therefore, a qualitative research design will be used using the interpretivism paradigm.

Pragmatism focuses on human experience seeking meaning from eliciting these. For the purposes of this study, this offers too narrow an interpretation to allow for existing works and influences to be fully considered as the individuals interviewed may not have personal experience of, or know of, such guidance. Critical inquiry is the pursuit of change with the aim of persuading others an argument is valid and this, is contended, is not a suitable approach with a subject like privacy where collaboration and understanding will be required for any changes to be implemented and take effect. Positivists confine themselves to absolute truth, a concept that cannot be proven (Popper 2005, Guba 1990), making this paradigm unsuitable for qualitative research (Guba 1990, Arghode 2012). While post-positivists acknowledge that absolute truth is not achievable, they believe there is ’a single truth’ (Popper 2005) whereas, interprevism acknowledges that multiple truths exist (Arghode 2012).

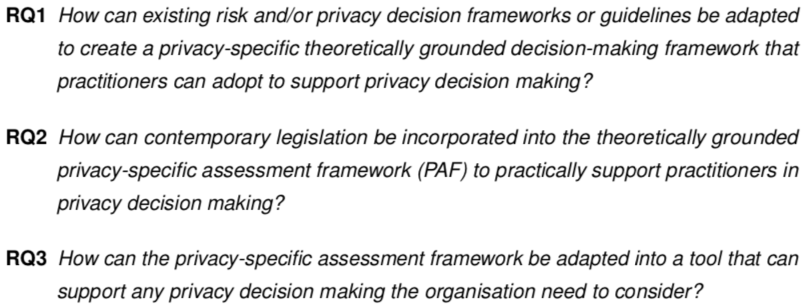
Interpretivism accentuates the importance of understanding rather than explaining (Blaxter et al. 2010). This allows for ’multiple truths’ to be captured, each contributing towards the generation of knowledge (Arghode 2012). Further, interpretivism supports utilising mixed and/or multiple methods and studies while recognises that validation is paramount to help explain and analyse the data and reduce bias during the study (Creswell 2009).

Research Questions and Propositions

The research questions (RQ) have been presented in the introduction (see Chapter 1, Section 1.2) however, for ease of reading, the main research question is repeated here:

The main question this work asks is:

> *”How can privacy assessment be incorporated into organisational decision-making in a practical manner, encompassing legal and contextual considerations, to provide repeatable, effective decision support for determining privacy risks and facilitating the integration of the privacy decision-making function into organisational decision-making by default?”*



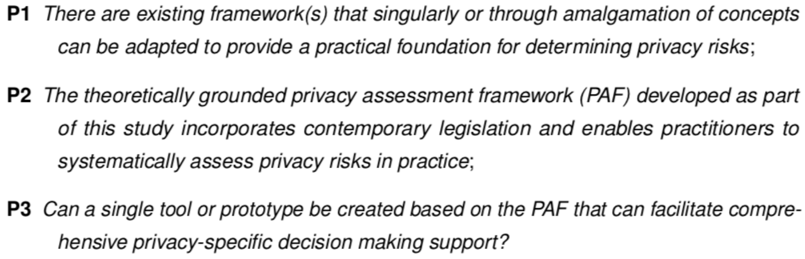
RQ1 *How can existing risk and/or privacy decision frameworks or guidelines be adapted to create a privacy-specific theoretically grounded decision-making framework that practitioners can adopt to support privacy decision making?*

RQ2 *How can contemporary legislation be incorporated into the theoretically grounded privacy-specific assessment framework (PAF) to practically support practitioners in privacy decision making?*

RQ3 *How can the privacy-specific assessment framework be adapted into a tool that can support any privacy decision making the organisation need to consider?*

Figure 3.1: Supporting Research Questions

This is further broken down into 3 sub-research questions (RQ), see Figure 3.1, and three underlying propositions (P), see Figure 3.2.



P1 *There are existing framework(s) that singularly or through amalgamation of concepts can be adapted to provide a practical foundation for determining privacy risks*;

P2 *The theoretically grounded privacy assessment framework (PAF) developed as part of this study incorporates contemporary legislation and enables practitioners to systematically assess privacy risks in practice*;

P3 *Can a single tool or prototype be created based on the PAF that can facilitate comprehensive privacy-specific decision making support?*

Figure 3.2: Supporting Propositions

Ensuring research conducted is valid, sound and defendable, forms an essential part of a study. For validity, it is important for the researcher to show that work has been conducted systematically, following an appropriate, recognised methodology because, the basis of producing knowledge within a given field is very much dependant on method (Pinsonneault and Kraemer 1993). In social research, to demonstrate the quality of empirical social research, including four types of quality or validity checks is recommended, these are: construct validity; internal validity, external validity and reliability (Yin 2013, Rowley 2002). The approach to each validity test is shown in Table 3.7.

Validity Test	Approach to be taken
Construct Validity i.e. how well do the questions asked measure the effectiveness of privacy decision-making?	This will be achieved through: First, establishing a *"chain of evidence"* by trialling existing questionnaires and/or frameworks; any questions developed based on interpretation and/or analysis of existing frameworks will be linking questions back to the literature, thereby reducing subjectivity. Second, using multiple sources of evidence, this will include existing documentation, interview transcripts and completed questionnaires.
Internal Validity i.e. showing a causal relationship between variables	data analysis
External Validity i.e. how generalisable or repeatable are the results of the study(ies)	This will be achieved by research design, ensuring theory informs the work and through thorough development and definition of method and logic applied as part of the study(ies).
Reliability i.e. how replicable is the study, will a repeat result in the same outcome and/or findings?	To be achieved thorough research design (case study protocol) and, as part of each study carried out, multiple trials/applications of the trial/questionnaire will be carried out which will aid in demonstrating replicability and reliability.

Table 3.7: Validity Checks (adapted from (Yin 2013, Rowley 2002)

3.5.2 Data Collection

Unit of Analysis

The unit of analysis that forms the basis of the case study is *the privacy practitioner(s)*. As the person who makes decisions around privacy, the privacy practitioner may not have the title of privacy practitioner in all organisations but it refers to the person(s) responsible for privacy decision making within an organisation. For the purposes of this work the privacy practitioner will be referred to merely as "the practitioner(s)".

Data Collection Approach

The literature review identified that public bodies have a legal obligation to publish data in open format. Therefore, public bodies do not have a choice in whether or not to make data available meaning public bodies should have a vested interest in ensuring privacy has been preserved prior to publication to avoid any potential data protection breaches and *public open data* was chosen as the domain in which the framework devised to answer RQ1 would be tested (see Sections 1.1.1 and 2.3.1. Further, because the whole idea of making data available in open format requires un-controlled and un-regulated access and use of the data, privacy preservation in this sphere is particularly difficult and therefore, it provides an excellent subject of study as it is vitally important that privacy has been preserved prior to publication.

Thus, to answer the research questions, a number of sub-cases will be conducted, these will consist of a combination of methods and approaches, starting with a preliminary scoping exercise (see Section 3.4.10) to establish current views and processes used for privacy decision-making within public bodies. This will take the form of a freedom of information request submitted to a randomly selected group of local authorities (LAs), asking to what extent they publish data in open format (see Section 4.2). The outcome of this scoping study will be used to inform whether any further scoping studies are needed to help establish how public body practitioners make privacy decisions for open data publishing (see Chapter 4).

The data collection will be guided by this protocol and consist initially of collecting data from public bodies, the literature and reports and publications relating to privacy and decision making published by recognised official bodies and institutions.

In addition to case study method, a number of supporting data collection techniques will be used to collect information. These are outlined in Sections 3.4.6 to 3.4.11.

Validation

In gathering evidence the following three research principles will be applied:

1. *Triangulation* i.e. evidence will be collected from multiple sources and this will be used to inform and/or corroborate and verify one or more findings or facts;

2. *Database of Evidence* i.e. collating the evidence gathered in a central database. For this purpose, SPSS and NVivo will be utilised to collate all evidence gathered as part of the case study.

3. *Chain of Evidence* This will be evidenced throughout the work by including appropriate documentation such as participant questionnaires etc. and using citations from the interviews and documents gathered (adapted from (Rowley 2002)).

3.5.3 Analysis and interpretation

Once data has been collected as part of a study, the data will be analysed and interpreted using a combination of interview, modelling and data coding techniques. The data coding will be done using SPSS and Nvivo software following recognised coding techniques such as grounded theory (see Section 3.4.3).

Data collected will be analysed and interpreted using modelling techniques, coding of data using techniques including; grounded theory (see Section 3.4.3), analysis and interpretation.

3.5.4 Manner of presentation

Each study undertaken as part of this case study will be conducted in accordance with method protocol. This means that, where appropriate, each study will include a protocol and/or study plan or schedule detailing processes to be followed. Data collected will be analysed and reported on as part of that study. Each study will then be presented as a separate chapter within the main body of work to form part of the overall case study.

This will allow for corroboration and verification of evidence and demonstrate the progress made as research progresses and more findings are made.

Chapter 4

Scoping Studies

4.1 Introduction

The question this first piece of work seeks to answer is how public bodies make decisions around privacy for open data publishing. The literature review confirmed that public bodies have a legal obligation to make data available in open format. Further, it also found that, although there are frameworks available that can be used to assist in making privacy decisions, these include privacy at a high-level only. Some guidelines, such as the NIST guide to PII (see Section 2.7.2) and BS ISO29100:2011 (see Section 2.7.3), consider privacy as part of other considerations, but essentially view privacy from a security perspective and how to wrap security around data within IT systems. Others, like PbD (see Section 2.7.4) and PIA (see Section 2.7.1) provide high-level guidance only on how to determine privacy risks. However, little is know about how the practitioners themselves see their obligation to publish in open format, or indeed whether or how they meet that requirement.

Therefore, this work will begin with a scoping study (see Chapter 3, Section 3.4.10), consisting of a freedom of information request to get some idea of how many local authorities currently have an open data portal and the extent to which they publish data in open format (Section 4.2). To validate the findings from this study, an analysis of how many local authorities in the UK had open data portals in 2016, and the extent to which they published data in open format through the data.gov.uk website will be conducted, the results of this can be found in Section 4.3.

Following on from these studies, a series of semi-structured interviews with public body practitioners will be conducted to ascertain whether they currently publish information in open format and how decisions are made around such publication. The interviews will also ask how decisions are made regarding privacy. These interviews will be recorded and transcribed, details of the outcomes of this study can be found in Section 4.4.

In parallel with this work, a thorough search of the literature and official sources such as the Information Commissioners Office (ICO) and other government bodies will be conducted to find out what guidance is available for practitioners, as well as any supplementary source that may provide assistance or guidance in open data publishing (Section 4.5).

The transcripts from the contextual interviews and any guiding documents found during the search will be uploaded into NVivo (a qualitative synthesis software application) and will be coded using Corbin and Strauss grounded theory model (Corbin and Strauss 2008). Because these pieces of research will be undertaken as a scoping exercise of work, they will be undertaken with a view that the findings will help frame the format of the rest of this study. Therefore, the coding exercise will be conducted to establish patterns and common terminology used in the open data publishing arena and to find out what current practices are being employed with regards to open data publishing. It will also be used to gather the thoughts and feelings of practitioners on the subject of open data publishing rather than to generate theory. However, grounded theory can, and commonly has been, used in information system studies as a coding technique, rather than as a method for generating data (Urquhart et al. 2010) which suits the purpose of this piece of work.

4.2 FOI Request

To gauge how proactive LAs in the UK have been in meeting their legal obligations in publishing information open source, a freedom of information request (FOIR, see Section 2.3.1) was submitted to 22 randomly selected local authorities (LA). This FOIR sought to establish whether they publish data in open format either through the freedom of information publication scheme or an open data portal as well as asking for contact details of the people responsible for publication. A copy of the request can be found in Appendix A.

4.2.1 FOI Responses

The FOI request asked; (i) whether the LA published open data; (ii) if the LA has an open data portal or website; (iii) whether they contribute open data to the national open data portal; (iv) the name and contact details for the person(s) responsible for open data publication; and (v) details of their job title or role is within the LA. All 22 LAs responded. The responses were entered into a spreadsheet and analysed. The findings were:

- All the LAs published data under the FOI publication scheme;
- Six of the LAs had an open data portal;

Thus, although all published something under the FOI publication scheme, only 37% had some form of open data platform or portal. Another finding was that the role and/or department responsible for open source publishing varied considerably across the LA's contacted, ranging from Information Officers through to Legal or GIS professionals.

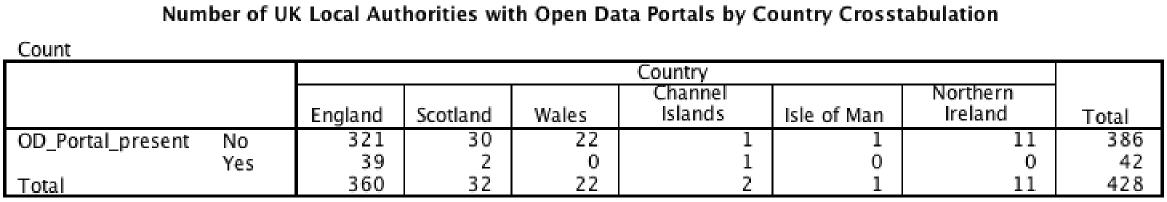
This suggested that not many LA's publish data that is not required to be published under the FOI publication scheme (see Section 2.3.1). However, the data published under the FOI publication scheme is not always in open format, rather it consists of links to PDF files or webpages where relevant information can be found. Data is only required to be published as open data if it has been requested under FOI and/or ROPSIR (Information Commissioners Office 2015b).

Therefore, a more in-depth study was carried out on LA publication habits to find out how many proactively publish data in open format without necessarily waiting for someone to request the data. To this end, a secondary dataset, created by Owen Boswarva, a member of data.gov.uk was utilised (Boswarva 2016). This dataset will be referred to as 'the Boswarva dataset'.

4.3 Local Authority Open Data Portals

The Boswarva dataset consists of a fusion table, i.e. an amalgamation of multiple data sources into one spreadsheet (Google 2015), containing a list of all LAs in the UK with links to open data portal where these were available. This data was collected from the data.gov.uk site under Open Government License (Boswarva 2016).

The data table was downloaded and coded and analysed using SPSS (a statistical analysis software package) with the following findings:



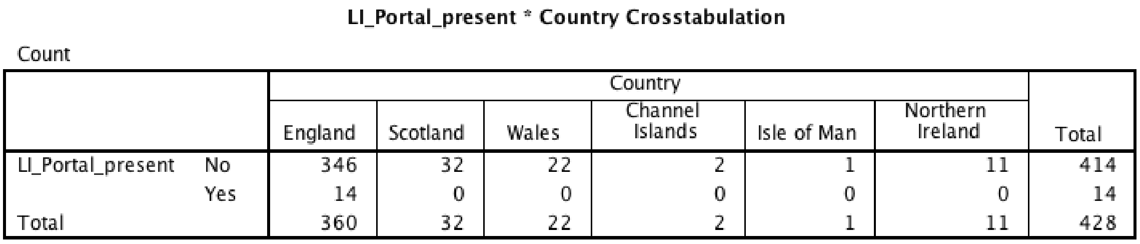
Number of UK Local Authorities with Open Data Portals by Country Crosstabulation

Count

		Country						Total
		England	Scotland	Wales	Channel Islands	Isle of Man	Northern Ireland	
OD_Portal_present	No	321	30	22	1	1	11	386
	Yes	39	2	0	1	0	0	42
Total		360	32	22	2	1	11	428

Figure 4.1: Number of UK LAs with Open Data Portals

There are 428 LAs in the UK, of these 42 have open data portals (see Figure 4.1). In addition, there are 14 LA's who submit data to a county or area combined open data portal (see Appendix A, Figure 4.2). Of the LAs who contribute to a combined open data portal, 3 also have LA specific open data portals. These figures show that only 12 percent of LAs across the whole of the UK contribute to either LA specific and/or combined open data portals indicating that the majority of LAs have chosen not to publish open data as standard as yet, despite any legal obligations to do so. This is a much smaller percentage than the sample group of LAs contacted as part of the FOI request, suggesting this sample group are not fully representative of all LAs in the UK.

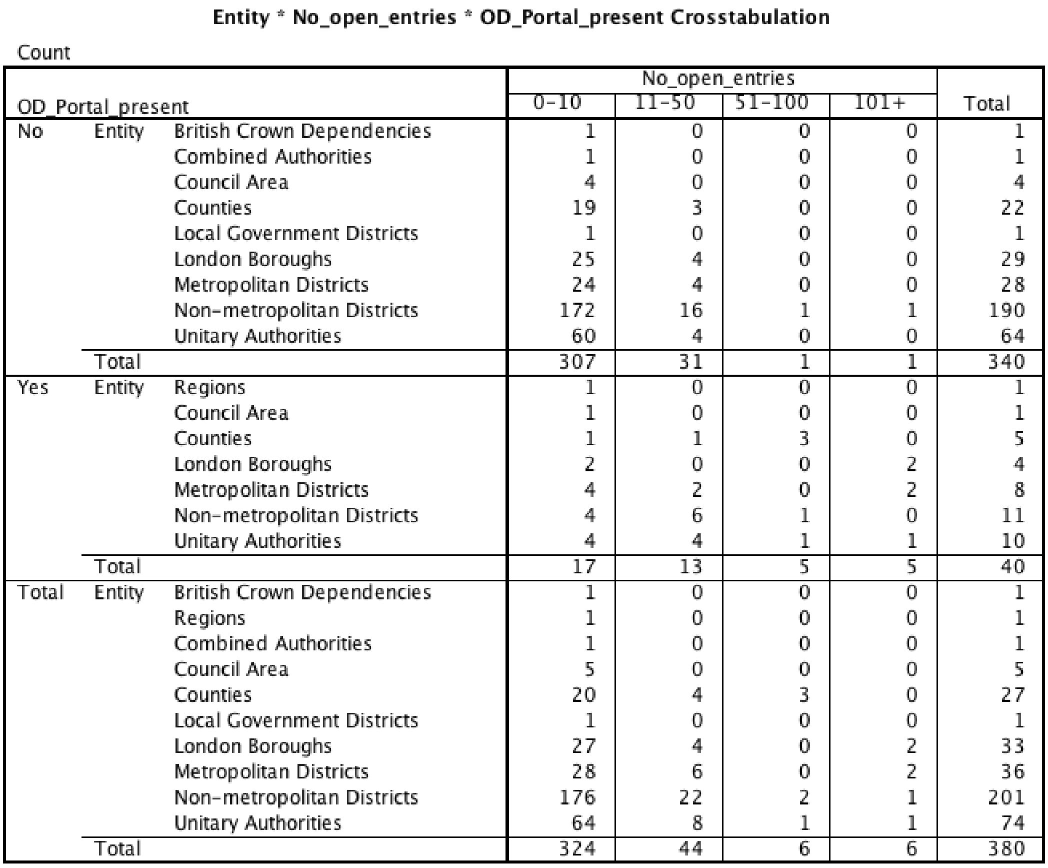


LI_Portal_present * Country Crosstabulation

Count

		Country						Total
		England	Scotland	Wales	Channel Islands	Isle of Man	Northern Ireland	
LI_Portal_present	No	346	32	22	2	1	11	414
	Yes	14	0	0	0	0	0	14
Total		360	32	22	2	1	11	428

Figure 4.2: Number of UK LAs with Combined Open Data Portals



Entity * No_open_entries * OD_Portal_present Crosstabulation

Count

OD_Portal_present			No_open_entries				Total
			0-10	11-50	51-100	101+	
No	Entity	British Crown Dependencies	1	0	0	0	1
		Combined Authorities	1	0	0	0	1
		Council Area	4	0	0	0	4
		Counties	19	3	0	0	22
		Local Government Districts	1	0	0	0	1
		London Boroughs	25	4	0	0	29
		Metropolitan Districts	24	4	0	0	28
		Non-metropolitan Districts	172	16	1	1	190
		Unitary Authorities	60	4	0	0	64
	Total		307	31	1	1	340
Yes	Entity	Regions	1	0	0	0	1
		Council Area	1	0	0	0	1
		Counties	1	1	3	0	5
		London Boroughs	2	0	0	2	4
		Metropolitan Districts	4	2	0	2	8
		Non-metropolitan Districts	4	6	1	0	11
		Unitary Authorities	4	4	1	1	10
	Total		17	13	5	5	40
Total	Entity	British Crown Dependencies	1	0	0	0	1
		Regions	1	0	0	0	1
		Combined Authorities	1	0	0	0	1
		Council Area	5	0	0	0	5
		Counties	20	4	3	0	27
		Local Government Districts	1	0	0	0	1
		London Boroughs	27	4	0	2	33
		Metropolitan Districts	28	6	0	2	36
		Non-metropolitan Districts	176	22	2	1	201
		Unitary Authorities	64	8	1	1	74
	Total		324	44	6	6	380

Figure 4.3: Number of open datasets published through data.gov.uk by LA type

Upon conducting a search of how many datasets were uploaded to the central open data government repository at data.gov.uk, in total 380 datasets were published on the portal (see Figure 4.3). Of those only 40 originated from a LA with an open data portal. Thus, looking at these figures, it appears many LAs contribute to this portal even if they do not have dedicated open data portals. That would indicate that many LAs publish data in open format using the central data.gov site rather than through dedicated open data portals, suggesting that the sample group of respondent LAs from the FOI scoping study may have included a higher proportion of LAs with an open data portal. Further, it appears that 2 of the LAs who have open data portals do not contribute to data.gov.uk. In view of that, perhaps using open data portal presence as an indicator for open data publication is not a suitable metric.

What these figures also reveal, is that, of the LAs who publish open data, 85 percent have published less than 10 datasets. This would indicate that, of the 12 percent who do publish data in open format, most do not do so on a regular basis. This is in line with the findings from the FOI request (see Section 4.2) which showed that most LAs do not publish data in open format proactively.

The results of these two studies (Sections 4.2 and 4.3) suggest that, unless the data is requested under FOI or ROPSIR, it will not be available. However, to establish whether that is the case, some further work will need to be conducted. To this end, a decision was made to make some more in-depth enquiries into whether that might be the case and gauge the opinion of practitioners within the LAs.

4.4 Local Authority Interviews

This scoping study is an explorative study of how public bodies currently make decision about data privacy in relation to open data publishing, it will be conducted through semi-structured interviews (see Section 3.4.6). The intention was that the interviews could be used to guide the study and establish how LAs currently deal with open data publishing and also, test the waters to see whether there would be sufficient interest in collaborating with the researcher in testing any prototype tool created as part of this work.

LA practitioners face a lot of constraints on their time and therefore, getting practitioners to agree to participate in research is not an easy task. Six of the LAs who responded to the FOI requests in Section 4.2, were contacted again by telephone and/or email to ask if they would be willing to participate in a semi-structured interviews (see Section 3.5.2). Two of these agreed to take part which, in view of the restrictions on practitioners time, was considered a successful outcome. Further, because these interviews were intended to help inform and guide the study, a small sample of interviews was considered sufficient to provide useful insight (de Ruyter and Scholl 1998).

Three interviews were conducted at the LAs offices with five practitioners participating from across the two LAs, two managers and three practitioners who worked with open data. The interviews were taped and transcribed and loaded into NVivo for analysis. The participants consisted of two members of staff from each LA. These interviewees worked in the following capacities; three were IT professionals who work with data and open data; one was a General Manager with responsibility, among other things, for data and FOI requests; and one was the Manager of an Open Data Department. The interviews took place at the offices of the LAs in three sessions. The first session was attended by two interviewees (an IT professional and the General Manager), lasting lasted 1.5 hours. The other two sessions were attended by the other two interviewees separately, lasting approximately 1 hour each.

As part of the conversation, it was explained that the outcome of these interviews would be used to gain a better understanding of their data publishing practices and what barriers to publication they may have encountered. The interview guide for these interviews can be found in Appendix A, Section A.1.

Prior to the interviews, the participants were sent brief details of the research being conducted, a list of questions and a participant agreement form (see Appendix A, Sections A.2 and A.3 for copies of these).

4.4.1 Coding

The interview transcripts were coded based on Corbin and Strauss (2008) grounded theory model. The first stage of coding in grounded theory is the open coding phase where snippets from the transcripts are grouped and broken down into named codes through examination, conceptualisation and comparing (Corbin and Strauss 2008). The coding was led by the information within the transcripts rather than by the questions asked. This was done deliberately so as not to miss any points made that might have arisen from the conversation flow rather than from a specific question. This exercise resulted in 77 codes being created.

In the second stage, the Axial coding stage (Corbin and Strauss (2008), see Section 3.4.3), these codes were amalgamated into 13 overarching categories by grouping them into related themes. The 13 themes were:

1. Aspirations - this category relates to the aspirations LA has for open data in future and what they would like to achieve;

2. Benefits - this category relates to what the practitioner(s) see as the perceived benefits of open data publishing for the LA;

3. Current processes - relates to how whether open data is currently published and, where applicable, the processes in place for publication;

4. Data management and manipulation - relates to how data is currently managed and/or published;

5. Departments - relates to the department and the team(s) involved in open data publication;

6. History - information relating to the history of open data publishing from the perspective of the LA;

7. Obstacles - relates to any real or perceived barriers to open data release;

8. Other LA Platforms - information about what other LAs are doing in the open data publishing field;
9. Policy - relates to any formal policies in place;
10. Privacy - this category includes privacy concerns, risks and current approach to privacy;
11. Publication Scheme - relates to releasing information under the FOI publication scheme;
12. Roadmap - what plans are in place for the future of open data publishing within the LA;
13. Transparency - any discussion about transparency and how this relates to open data publishing.

No further coding took place of this data because the coding was not intended to be used for generating theory in the traditional sense of grounded theory, where the third phase would be to carry out selective coding to generate theory (Corbin and Strauss 2008). Rather, the coding exercise was done to gain a better understanding of the underlying messages and make sense of the interviews.

Relating these categories to the questions asked the following findings became evident:

Q1 - Team/Department

One LA stated they originally had 40 people scattered around the authority who were working with data and, as part of the open data initiative within the authority, *"that bringing them together in one space would be good" (P1)*. However, it also transpired that, of those, only one person worked full time on open data publication with a second assisting on an ad-hoc basis. The other LA did not have a team in place rather they stated that their IT Information Governance Officer was the only person who did any work in this area, and that involved mainly ensuring that the LA's obligation under the FOI publication scheme were met.

Q2 - Policy

Neither LA had any formal policies in place for publishing data in open format, one practitioner stated: *"it is easy enough to say, yes, lets adopt a policy but then you find nothing has happened" (P3)*. The reason for this may be that getting agreement on what should be included and creating formal policies in this area might be considered too difficult, as eluded to by another practitioner who stated they decided to: *"just get going on releasing data but not get bogged down and do anything that was too hard" (P1)*.

Q3 - Process

When asked about processes one practitioner stated: *”we rely on the process maps...and they are quite specific for creating/updating particular types of datasets” (P2)*. Thus, through the interviews it became quite apparent that few formal processes were in place for what data to publish or indeed how to publish open data, one practitioner stated: *”it has been more driven by trying to reduce FOI request numbers rather than trying to be proactive with the data....we sort of free wheeled” (P4)*. This is perhaps not surprising given that no policies are in place for this area either. When asked why they thought there were so few formal processes in place, one practitioner stated it would be good to *”put it into a good standard, keep it updated and take ownership responsibility for it. That’s the difficult bit” (P2)*.

Q4 - Data

From the interviews it became evident that one of the LAs practice was not to publish until pressure dictates otherwise: *”why do we have to do anything when we can get away with the bare minimum?” (P3)*. In contrast, the other LA published open data regularly with performance indicators (processed statistical information only) representing around 80 percent of output, *”we are looking at a massive number of indicators, 1000’s and 1000’s of indicators” (P2)* . The publishing of these indicators was predominantly automated. When asked how many raw datasets have been published, the practitioner stated: *”outside of the performance stuff, I think if you got up to 100, you would be doing fantastically well...its 80 percent of our time that is spent on that 20 percent of the data” (P5)*. Therefore, although the LA publishes data in open format, most of this is performance indicators produced as part of reporting processes for different areas of the business. When asked what the reason for this low number of raw datasets being published when compared to performance indicators, the reply was: *”It’s economy of scale” (P1)*.

Q5 - Problems/obstacles

There were many barriers raised for why data could not be published in open format. These included:

- **Lack of resource** - *”the issue has always been the cost of releasing that data and the difficulty in doing it on a repetitive basis” (P1)*;
- **Culture** - *”When you are trying to get service areas to try and publish data that they would much rather be kept secret, just because it is culture” (P4)*;
- **Technical challenges** - *”Data is spread across a lot of different systems...it is difficult in some cases to get that data out of the system” (P5)*, *”It’s like an algorithm that*

fights back at you" (P2);

- **Lack of collaboration** - *"Not everyone wants to collaborate" (P3)*;
- **Management buy in** - *"what we have got to do is get buy in and the attitude is, well, what's in it for us?" (P3)*;
- **Meeting user expectations** - *"I don't think we could ever really meet the ambition of the market and what they would want us to put out" (P4)*.

Q6 - Standard

Although the practitioners interviewed all agreed that having a standard for open data publication would be useful, they felt that the technical barriers such as legacy IT systems and the difficulties in getting data output from these in consistent formats would prevent this from being a realistic goal for now. However, they would all be happy to participate in developing such a standard if a group was set up to achieve this. For the moment, the best hope, according to one practitioner, would be to create standardised schemas: *"there are attempts towards standardising the actual schemas which will be absolutely brilliant. It is never going to work 100 percent but even if it works 50 percent it will still be brilliant" (P3)*.

4.4.2 Other findings from interviews

When asked about privacy, practitioner opinions ranged from cautious: *"some information, particularly datasets containing sensitive personal data, will clearly present a need for caution, and the anonymisation issues may be complex for large datasets containing a wide range of personal data" (P4)"* to unconcerned: *"I am not too concerned about the privacy angle because we would never put out personalised data." (P1)*, and very concerned: *"I am almost convinced that if I went back through our data that we have published over the last 4-5 years, I would find something that we'd missed" (P2)*. This may explain why only 6 of the LAs contacted as part of the FOIR currently have an open data portal.

Where public open data is published, innovative uses have been made of the data. For example, OpenStreetMap used public open data to provide users with free access to maps of their local area (Open Data Institute 2015). However, while the LAs interviewed appreciated and appeared to endorse government's eagerness to promote open data publication as a means of increasing and promoting transparency (see Section 2.3), e.g. *"by opening up information to people you can foster growth" (P1)*, most do not practice what they preach.

These findings suggest that there is great disparity in how each of these LAs approach open data publishing. One of the LAs confirm the finding from the two previous studies into LAs open data publishing approaches that, unless the information is requested, it is unlikely to be made available (see Sections 4.2 and 4.3), the other LA interviewed however, do publish open data proactively. However, looking at the findings from the Boswarva dataset, only six of the LAs have published in excess of 50 datasets on the central data.gov.uk site (see Figure 4.3), suggesting this LA is in the minority.

4.5 Open Data Publication Guidance for Practitioners

In parallel with the interviews a search was conducted to establish what assistance and guidance is available to practitioners who are required to publish data in open format. This search concentrated on legal obligations to publish and the guidance available surrounding that with regards to what data to publish, how to publish the data, required format etc.

The search resulted in a series documents being selected for coding, these and the reasons for why those documents were chosen are listed in Appendix A, Section A.4, Table A.2.

Like the interviews, these guidance documents were loaded into NVivo and coded following Corbin and Strauss (2008) (see Section 3.4.3). The open coding phase resulted in 118 codes. This categorisation of the documents was used to inform the next stage of coding (*Axial coding*, resulting in the open codes being reorganised and grouped (categorised) into 8 overarching categories based on the advice or guidance they provided, brief descriptions of the type of information contained within each of these categories can be found in Appendix A, Section A.4 (see also Figure 4.5).

What also became evident as part of the axial coding was that these guidance documents could be separated into three distinct types of guidance or advice as follows:

1. Overarching advice i.e. general advice about open data, what it is and governance around open data publishing;
2. Preparation i.e. advise on how to prepare data for open data publication;
3. Specific i.e. advice on specific legislation and how to meet obligations under that legislation.

With regards to documentation, Figure 4.4 show which documents are considered to fall into each of these categories of advice.

The outcome of this exercise showed quite clearly that, although some of these guidance documents provide more detailed advice on the how, what, when, where and who, this was confined to the third overarching type of advice which provided guidance about specific pieces of legislation, the FOI and ROPSIR. Further, while these documents provided a bit more detail about the types of data that might be published within each category, this advice consisted of suggestions rather than clear direction for how to

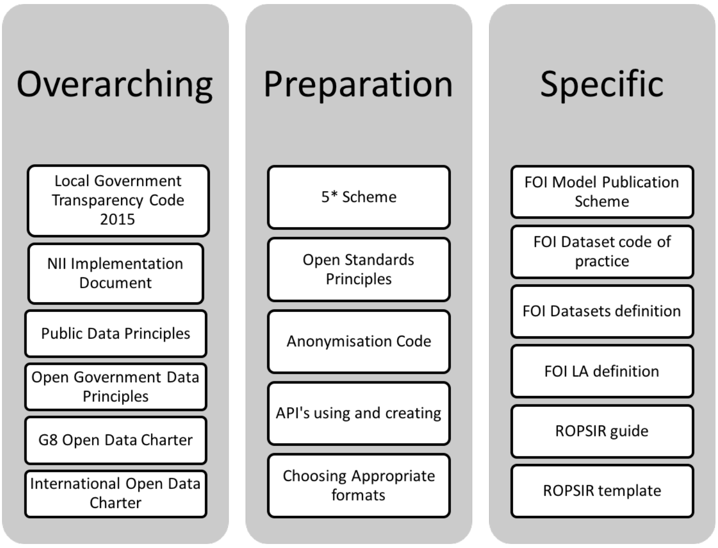

Overarching
Local Government Transparency Code 2015
NII Implementation Document
Public Data Principles
Open Government Data Principles
G8 Open Data Charter
International Open Data Charter
Preparation
5* Scheme
Open Standards Principles
Anonymisation Code
API's using and creating
Choosing Appropriate formats
Specific
FOI Model Publication Scheme
FOI Dataset code of practice
FOI Datasets definition
FOI LA definition
ROPSIR guide
ROPSIR template

Figure 4.4: Guidance documents in each category

implement the schemes. The rest of the guidance documents provided only high-level and non-specific advice. Moreover, most of this can be related back to the Government strategy of being transparent, i.e. what transparency means; what it is attempting to achieve as a government, and therefore public body, objective and how open data is one of the tools that will help meet that objective.

The third stage of coding, the selective coding (see Section 3.4.3), was done by looking at these codes from a business perspective, through the lens of the practitioner. From this perspective, it can be argued that these categories should be viewed and discussed in terms of how they fit into an organisation's strategic logic (Sanchez and Heene 2004). This then means that the codes can be grouped into the relevant strategic sphere or themes as follows:

Strategy i.e. what are we aspiring to achieve;

Objectives i.e. what do we think we must we do to achieve the strategy;

Actions i.e. what tasks must we preform to achieve the objectives.

Thus, following this logic, strategy, objectives and action would form the overall strategic themes and the axial codes would fit in beneath that as depicted in Figure 4.5.

4.6 Conclusion

This Chapter has outlined the findings from the following scoping studies:

1. A freedom of information request was sent to all local authorities (LAs) in a randomly selected county in the UK, to establish whether they currently publish information open source (see Section 4.2);

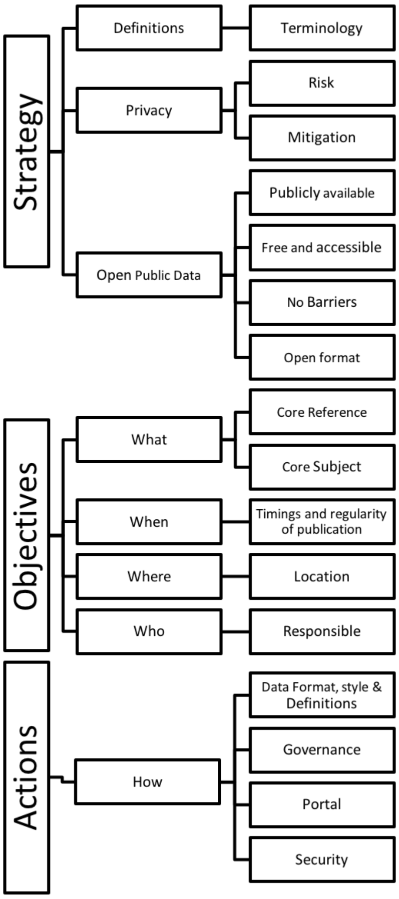


Figure 4.5: Strategic Themes & Axial Codes

2. Validation of the findings from the FOI requests through a review of existing guidance on what data public bodies should publish open source (see Section 4.3);

3. Five contextual interviews were conducted public body employees to find out about current practices in open source publishing (see Section 4.4);

4. A review of existing guidance available for practitioners to follow in Section 4.5.

4.6.1 Summary of Scoping Study Findings

The results of the interviews indicated that;

Making data available - While most LAs have publish something, this predominantly covers requirements under the FOI publication scheme.

Publishing Open Source - The majority of LAs do not publish datasets in open format unless it has been specifically requested (e.g. under ROPSIR or FOI);

Guidance - Existing guidelines provide no guidance on privacy decision-making;

Decision Making - None of the public bodies contacted have a formal decision making process in place around privacy decision making.

LAs are reluctant to publish open data proactively, those that do publish data choose to publish performance indicators that the LA are obliged to prepare as part of their standard reporting procedures rather than complete raw datasets. Thus, it appears that publishing the raw data poses a much larger problem than publishing the performance indicators and, as a result, the LA is aiming for quantity rather than quality, perhaps to satisfy a perception of proactively publishing data in open format.

Looking at the result of coding the guidance available to practitioners from government bodies such as the ICO, these show that while guidance documents exist, most guidance documents provide limited assistance with only high-level and non-specific guidance. The coding of the guidance documents showed how these can be grouped firstly, by type of advice (see Figure 4.4) and secondly, by strategic logic into themes (see Figure 4.5).

What the final themes that emerged from this exercise clearly showed was that, of the guidance documents available, while there is a lot of strategic advice and a fair amount of advice relating to objectives, there is very little action advice (see Figure 4.5). Further, none of the documents provided any detailed advice or guidance on how to deal with privacy risks in making data available in open format other than to ensure DPA obligations are met. Thus, although there is a lot of guidance documents available, these do not provide much practical assistance to practitioners in regards to making informed decisions about privacy risks.

Similarly, the literature review confirmed that, while there are frameworks and guidance documents that can be used to support privacy decision-making, none provide detailed practical assistance for practitioners to follow when making decisions about privacy risks for open data publication (see Chapter 2, Section 2.8). This lack of ’usable’ guidance in relation to privacy leaves public body officials with little direction for how best to proceed when tackling open source publishing of their information.

4.6.2 Next Steps

The research question asks how privacy can be assessed and practicably incorporated into organisational decision making by default (see Section 1.2). The results of these scoping studies show that there is a real need for practical guidance on how to make decisions about privacy that can provide adequate assurances that privacy has been preserved prior to publication. Further, these studies have confirmed that, in order for LAs (and other public bodies) to become more proactive in publishing data in open format, they need to have formal processes in place to safeguard privacy before publication occurs. Thus, it is contended that having a framework and guidance documentation in place will provide a starting point for opening up more opportunities for proactive open data publishing.

RQ1 asks for consideration to be given to whether there are any existing frameworks that can be adapted to create a privacy-specific framework that can be applied in practice and encourage practitioners to make informed, transparent privacy decisions (see Section 1.2).

Contextual Integrity (CI) was identified as part of the literature review as being suitable for adaptation to meet this criteria Contextual integrity was developed by Nissenbaum (2010). It considers privacy in more depth than the other frameworks reviewed. CI asks that privacy is considered in light of both the data itself and the context within which the data was collected and it is contended this framework will provide a good starting point for privacy decision making in the open data publishing domain.

Based on the findings from the literature review and these scoping studies, further research will be conducted into how contextual integrity can be adapted and to support effective decision-making for assessing privacy risks for open data publishing. To this end, the first step will involve breaking CI down into its component elements and creating a meta-model to establish how best to achieve that goal.

Chapter 5

Conceptualising Contextual Integrity

5.1 Introduction

Chapter 4 concluded that public bodies do not have existing processes in place for making decisions about privacy before data is published in open format. However, irrespective of how the data is disseminated, organisations must determine whether and how they share or publish data and, to do this safely, they must, as part of this decision, consider the privacy of the individuals whose data they collect.

This work contends that because privacy is a broad concept, to truly assess privacy, a thorough assessment of the privacy risks and potential impacts of any breach must be conducted. Such assessment however, must go beyond viewing technical security, access controls and privacy laws. Rather, it should encompass a more holistic assessment at a strategic level, considering the dataset as a whole. This will require for example, that the context in which the data was collected, controlled and created also be taken into account within any decision on whether a particular dataset poses a privacy risk if published.

P1 hypothesises that: *”there are existing framework(s) that singularly or through amalgamation of concepts can be adapted to provide a practical foundation for determining privacy risks”* (Section 1.2).

The outcome of the scoping studies in Chapter 4, and the literature review (Chapter 2) established that adapting Nissenbaum’s Contextual Integrity framework (CI) (see Section 2.7.7) would be a suitable option for adaptation into a *privacy-specific assessment framework to practically support practitioners in privacy decision making*, (P1, see Section 1.2.1).

The problem is that CI is a theoretical framework that includes a number of interlinking and interdependent concepts. One way to make sense of complicated relationships and interdependencies is to visualise or model these using conceptual modelling techniques

(Faily and Fléchais 2010a). Within business, the idea of using modelling techniques to conceptualise risks has been used for decades (Mitchell and Nygaard 1999). For example, techniques commonly used in software design such as flow charting, system analysis or software system analysis diagrams have been used for risk identification (Province of British Columbia 2012) and for risk elicitation within, for example, security (Faily et al. 2012).

Therefore, to make sense of CI and to tease out how these concepts and ideas are related, the first step will be to create a meta-model of CI that can help readers visualise how the CI concepts relate and interlink. The intention is that this model will inform how CI might be applied in practice for assessing privacy risks and making informed privacy decisions around open data publication within a public body.

This rest of this chapter is organised in the following order. Starting with an explanation of the rationale for creating a meta-model in Section 5.2, the chapter continues with an overview of CI and how this has been interpreted to create a meta-model in Section 5.3. Next, in Section 5.4, each of the phases of the meta-model are explained and modelled, starting with *Explanation* in phase 1 (Section 5.4.1); followed by *Risk Assessment* in phase 2 (Section 5.4.2) and concluding with *Decision* in phase three (Section 5.4.3). This is followed by a worked example in Section 5.5, to illustrate how the meta-model can be used to support strategic privacy risk decision making in hypothetical public body, a public library. Section 5.6, concludes the chapter and outlines the next steps.

5.2 Rationale for creating a Meta-Model

Previous research using CI has predominantly been defined and discussed in theory (see Chapter 2, Section 2.7.7). Where CI has been applied in practice, it has been applied retrospectively to a specific problem or domain (Conley et al. 2012) or it has been applied as part of a system design process (Barth et al. 2006, Krupa and Vercouter 2012). In doing so, what these studies have done is to apply CI at a granular level, rather than holistically as part of an overarching privacy decision-making process. Thus, although previous work illustrates how the CI framework can be applied, these studies are insufficient for guiding decision-making in open data publishing.

In this domain, there are a number of constraints and unique considerations that need taking into account. First, the data being assessed will be existing data that has been created as part of a function of the public body unrelated to open data publication. Therefore, any decisions about what data and how this data has been collected, processed and stored will be historically pre-determined, and the decision-maker will most likely not have been involved in decisions around these factors. Second, for those datasets it will not be possible to fully define all elements because the role of the recipient cannot be specifically defined; anyone who downloads the data will be a recipient, and the data, once published, will be available to everyone. Creating a meta-model of CI will inform how the CI principles can potentially be applied to existing datasets in view of these constraints.

As part of CI, Nissenbaum asks readers to evaluate a particular information practice or flow against the prevailing context in light of social, political and ethical norms and values (see Section 2.7.7). However, Nissenbaum discusses these concepts in the abstract without providing any concrete structure as to how this might be achieved.

Creating a meta-model will serve as a visual aid to depict CI at a glance, thereby illustrating how a staged approach to privacy risk assessment can assist in making informed privacy decisions. Moreover, modelling CI can also be used to inform the development of potential privacy risk decision-making tool support that may help automate and streamline some of this process for practitioners going forward. This meta-model attempts to elicit the same outcomes as Nissenbaum's theory but illustrate this in a more tangible and structured format that will guide practitioners to the end point of making a decision.

5.3 Creating the Meta-Model

In terms of Nissenbaum, assessing privacy involves explanation, evaluation and prescription of the data, the actors and the transmission principles in context. Nissenbaum intended the framework as *"a descriptive tool"* rather than a prescribed methodology for applying the framework. In explaining the framework, she explains how the framework works by providing a lot of detail about the concepts and what each concept means without really setting out a step-by-step framework for readers to apply. Nissenbaum then goes on to validate the framework by providing worked examples for each concept. This model is the researchers interpretation of how these concepts can be translated into a visual, logical step-by-step process that can then be used to create a working framework for privacy decision-making.

5.3.1 Visual Representation

Unified Modelling Language (UML) diagramming is a universal visual language that is used to capture and represent concepts and the relationships between them (Rumbaugh et al. 2004). Thus, in the explanations that follow UML diagrams have been used to better illustrate each phase and provide readers with an easy point of reference and a better overview of how the elements relate within a phase (Fowler 2004).

5.3.2 Validation of Meta-Model

For validation the model will be aligned with Nissenbaum's validation approach as explained above combined with aligning the model to the strategic themes created in Chapter

4, Section 4.5, Figure 4.5. This will involve providing worked examples for each concept as the model is conceptualised to illustrate how this might be applied by a public body. This will take the form of applying the concepts to a theoretical public body, 'PB', wishing to assess the privacy risk of publishing a dataset in open format and relating this to the strategic themes and codes derived from the analysis of the guidance documentation conducted in Chapter 4 (see Figure 4.5). Then, to demonstrate how the meta-model might be applied in practice, a complete worked example will be provided in Section 5.5, applying the meta-model to a hypothetical PB, a library.

5.3.3 CI Guiding Principles

Within the explanation of how the framework works, two guiding sets of principles have been identified as being core to applying CI in practice:

1. The three key elements: explanation, evaluation and prescription (page 190);
2. The nine decision heuristics (page 182).

Thus, these principles have been used as the basis for creating a working meta-model of CI, dealing with each principle in turn.

5.3.4 Key Elements

The meta-model is intended as a practical, useful tool that will guide practitioners through CI and the privacy assessment process, culminating in a decision, to publish or not to publish.

In her book, Nissenbaum explains that the framework has three key elements; explanation, evaluation and prescription (p. 190). These elements provided a logical group of overarching categories that would frame the meta-model into understandable, logical progression steps. To this end, it is necessary to be clear about what information is being gathered (explanation) and assessed (evaluated) in order to make a decision (prescription) so that concrete actions and processes can be provided for the practitioners to follow.

In the meta-model these phases have been renamed; Explanation, Risk Assessment and Decision (see Figure 5.1) to better align with terminology that practitioners will relate to and to aid the flow of the staged approach that the meta-model will be asking practitioners to follow. The reasoning behind this change in terminology is discussed in more detail in Sections 5.4.2 and 5.4.3.

5.3.5 Decision Heuristics

Beneath the key elements (Phases), Nissenbaum proposes 9 decision heuristics (DH) that should be considered in relation to both existing and proposed new information flows. This

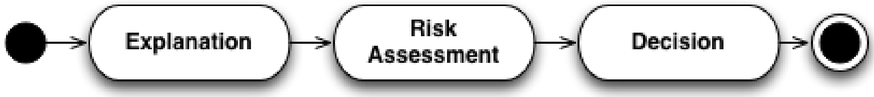


Figure 5.1: Meta-Model - Phases

will help establish whether privacy is likely to be, or has been, breached by a proposed new flow of information. These decision heuristics have been used to expand on the detail of information to be captured within each phase. The nine decision heuristics are described by Nissenbaum as follows:

1. ”Describe the proposed new practice in terms of information flows.
2. Identify the existing context.
3. Identify the actors, i.e. the information subjects, senders, and recipients.
4. Identify the transmission principles.
5. Describe existing entrenched informational norms and any significant points of departure.
6. Prima facie assessment.
7. Evaluation I: Establish whether there are any political or moral factors that could support/discourage publication e.g. what might be the potential effects or implications for justice, power structures, democracy etc.
8. Evaluation II: Establish how the system or practices could directly impinge on any values, goals, and ends of the context.
9. On the basis of these findings, contextual integrity recommends for or against the proposed new practices” (Nissenbaum 2010).

5.4 Meta-model

A public body will have a number of datasets that they need to assess for privacy risks. To do so they need to consider each dataset separately to establish what risks might be associated with publishing that dataset in open format. The meta-model asks that these risks are considered in stages (phases) to ensure that all risk are captured in light of the prevailing and future contexts. The first phase therefore, captures what data is included in the dataset, who have been involved in handling the data and what the prevailing context of the data collection and processing is. This is captured in phase one, *Explanation*.

5.4.1 Phase 1 - Explanation

The explanation element refers to the practice or system to be assessed. These should be assessed in view of any *"context-relative informational norms"* that may be breached. This should include an assessment of the key *"actors"*, i.e. the people that are/could be affected and their *"roles"*, as *"data subjects; data senders or data recipients"* It should also consider the *"attributes"*, i.e. the information itself (the data) and how this information is transmitted (*"transmission principles"*) and whether any changes to these elements potentially violate the existing or proposed new information flow.

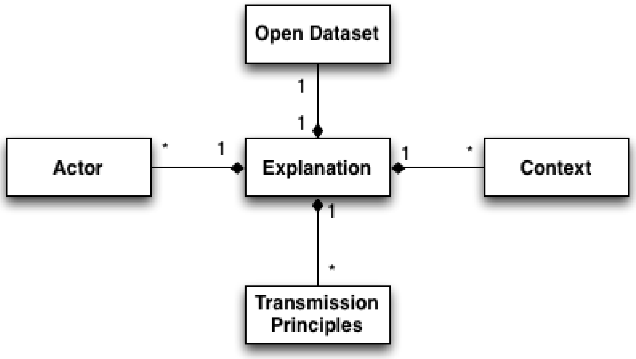


Figure 5.2: Explanation - Meta-Model

Explanation is the governing context used to determine whether any fundamental roles have or will be affected by changes in transmission principles. Thus, the explanation element forms the first phase of the meta-model. This phase involves establishing details about the dataset itself, the actors that handle the data or that the data is about, and the context in which the data has been collected, shared and transmitted. In this phase, information about existing informational norms is also collected.

The relationship between the explanation and the composite classes is one to many as there are multiple actors, contexts and transmission principles. However, the open dataset under review will have a one to one relationship with the explanation as there will only be one dataset under review for each explanation (see Figure 5.2).

To incorporate the decision heuristics (DH), it was determined that the first four DH's relate to gathering a more detailed overview of the data; the people (actors); the existing informational norms and transmission principles. The first decision heuristic (DH1), concerns the data itself and how it is proposed the data is to be transmitted. The second asks us to consider the existing context of the situation and environment surrounding the data and the people involved (DH2). The third, concerns the people involved with the data (DH3); and the fourth, seeks to establish how the data is currently transmitted (DH4). Thus, these four DH's were used to depict the explanation elements and how they relate.

At a more detailed level, explanation therefore requires that details are collected about the data itself, the people involved and their roles, how the data is transmitted and the context. Thus, the explanation is the superclass with each of the elements below depicted as subclasses (see Figure 5.3).

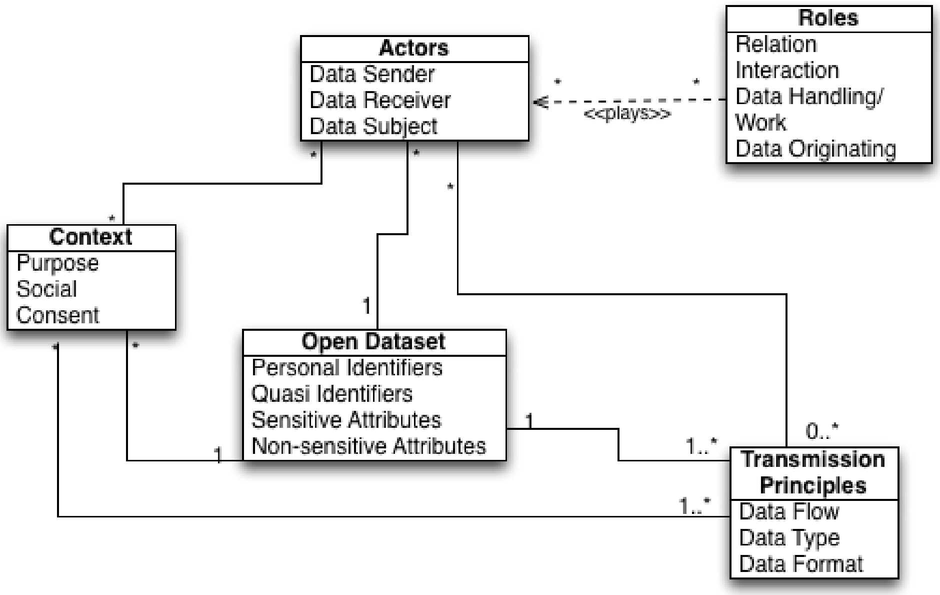


Figure 5.3: Explanation - Class Relationships

Figure 5.3 shows the relationship between the subclasses. These relationships can be explained as follows:

Open Dataset Each dataset needs to be considered separately to avoid overcomplicating the decision-making process and ensure all elements are thoroughly considered and CI is maintained throughout the assessment of each dataset. If we look at relations between the dataset and the other subclasses, there will always only be one dataset. Thus, the dataset will always be one in relation to each of the other subclasses. The attributes within the open dataset have been grouped by attribute type as follows:

Personal Identifiers (PI) i.e. directly identifying personal information;

Quasi Identifiers (QI) i.e. attributes that could, if linked with other QI's create a PI and thus, allow re-identification to occur (Henriksen-Bulmer and Jeary 2016);

Sensitive Attributes (SA) i.e. attributes that contain person specific but non-identifying information, such as disease or salary;

Non-sensitive Attributes (NA) i.e. attributes that have no directly or indirectly PI information (see Section 2.4.3 for more detailed descriptions).

Actors Each actor will act in one or more capacities. At data level, the actor will perform a data transmission role, sender, receiver or subject of the data being transmitted. It is also possible that the sender or the data subject may also download the data and thus, also become the receiver. Beyond the data transmission role however, the actor will also perform multiple relationship and/or work roles. For example, some actors will collate the data and may also know one or more of the data subjects. These actors may be one and the same or they may be different. Therefore, to allow for these nuances to be taken into account, the roles have been separated out as a class of its own.

Roles Any actor will act in the capacity of one or more roles in relation to the data and within each role the actor will have had different inputs. Therefore, each actor will be associated with multiple roles. The role may be context based such as how the actor is related to or interacts with another role or, it may be function based and defined by the work roles or duties. Thus, for this subclass, the relationship between the actor and each of his/her roles will be many to many.

Relation This group will contain details on what relationship the actors have to other actors, depicting both personal and professional relationships. For example, professional, family etc.

Interaction These attributes will contain information about how the actor interacts with the other actors such as citizen to professional or friend to citizen;

Data Handling this will capture information about what input or output the actor has in handling the data. They may have handled or processed the data or they may be the data controller who makes decisions around how the data is transmitted;

Work This refers to the occupation of the actor;

Data Originating This group seeks to capture information about the role of the data originator, it is possible the data has come from a third party and thus, the role of that third party needs to be considered as well.

Transmission Principles The transmission principles govern how information flows between actors. There are two sets of transmission principles, the existing and the proposed new flow of information, these are depicted as the data flow. In addition the data type and format may influence how the data can be transmitted and therefore, these have been added as considerations in this category as well. Whilst there will always be at least one transmission principle applicable when considering the subclasses of open dataset and context, for the actors subclass there may be no transmission principle applicable as not all actors will be involved in transmitting the data. Thus, the relationship between the transmission principles to actors will be zero or more to many.

Context There are many contexts that require consideration in relation to the other subclasses. Therefore, the relationship between this class and the other classes will be multiple to multiple. In the explanation phase, what needs to be captured for the full context to be considered in phase two, risk assessment, are, what Nissenbaum refers to as the; *"prevailing context"* (page 182). Therefore, in this phase, it is necessary to collect information about the context of how and why the data was captured in the first place. These contexts are:

Purpose This seeks to capture the original purpose of why the data was collected;

Social Capturing the social context in which the data was collected. For example, the data might have been collected from a school and thus, in an educational context;

Consent Where a dataset contains personal data or potentially sensitive data, the issue of consent will also need to be considered. Consent traditionally is considered from a legal perspective, discussing whether or not it is valid. However, consent also needs to be considered in relation to the data. Thus, in the meta-model at the explanation phase, what needs to be established is whether consent has been given and, if so, to what extent?

Validation - Explanation

In terms of the hypothetical PB and the strategic themes, the explanation phase is where they will record the dataset objectives of; 'what', 'who' and 'where' and part of the 'how' from the actions (see Figure 4.5) to include:

1. *What* - here the PB will outline details about what attributes types are in the dataset to be assessed and what information each of these contain (see Section 2.4.3);

2. *Who* - this will involve identifying the people within PB that are or have been involved in handling the data (the actors) and what role(s) they played in interacting with the data (data controller, processor and or subject, see Appendix 6.2, Section 6.3.2) For instance, was the data collected by a data controller from an internal department within the PB or by a data controller from a partner organisation?;

3. *Where* - here information about where the data was originally collected will be recorded (see Appendix 6.2, Section 6.3.2);

4. *How* - this involves recording how the data currently flows between stakeholders and what is the proposed new flow will be if the data is released as open data (transmission principles). Also, in what context was the data collected for each actor, for example, did one of the data processors (data sender) know the data subject (data subject) personally (governance).

Once PB has captured details of the data, actors, roles, transmission principles and the prevailing context, risks can be identified in light of legal obligations, established norms and values. To this end the LA will need to determine how publishing the data in open format might affect the transmission flow and what privacy risks might be associated with this new flow of data. This is captured in phase two, the risk assessment.

5.4.2 Phase 2 - Risk Assessment

The second element, *Evaluation*, relates to recognising the proposed change in transmission principles and how this will affect privacy in light of established norms, values and goals. Transmission principles refer to how the information is transmitted (information flow) at the moment and the proposed new information flow. Therefore, it is contended that what Nissenbaum is attempting to elicit in the evaluation element is the privacy risks associated with the proposed new information flows and how the changes in information flows might alter the values, norms and goals of the people involved (*"the actors"*).

In the meta-model this has been renamed 'Risk Assessment' because it is unlikely that Practitioners will be familiar with the term evaluation in this context. This change in terminology has been made for the following reasons:

Definition clarity Evaluate means to; *"assess or form an idea of the amount, number, or value of [something]"* whilst a 'risk assessment' is; *"A systematic process of evaluating the potential risks that may be involved in a projected activity or undertaking"* (Oxford Dictionaries 2016). Thus, both terms have a similar meaning. What the meta-model is seeking to elicit is the privacy risk associated with publishing data open source and therefore, risk assessment was chosen as a better definition of the desired outcomes for the meta-model;

Industry practice ISO/IEC29100, an internationally recognised privacy risk assessment framework, evaluates privacy in terms of risk. Further, practitioners will be accustomed to conducting risk assessments as part of their daily work. Thus, practitioners will be conversant with risk assessments and conducting risk assessments as part of their daily work and therefore, this terminology was felt to better resonate with them.

Therefore, for the purposes of the meta-model, the evaluation has been named risk assessment as it was felt that practitioners would relate better to that description. The risk assessment, consists of an evaluation of any privacy risks associated with a particular practice or transmission within a given context, taking into account how the information is conveyed or shared and the actors involved with that practice or transmission. Effectively, what the risk assessment is trying to achieve is an evaluation of the risks associated with any proposed changes or alterations in the data flow. In practice, practitioners are likely to be familiar with managing risk and conducting risk assessments as part of their duties, considering all manner of risk from security risk through to managing financial risks (see Section 2.6).

Risk assessment relates to recognising the proposed change in transmission principles and how this will affect privacy in light of established norms, values and goals. Transmission principles refer to how the information is transmitted (information flow) at the moment and the proposed new information flow. Therefore, it is contended that what Nissenbaum is attempting to elicit in the evaluation element is the privacy risks associated with the proposed new information flows and how the changes in information flows might alter the values, norms and goals of the people involved (*"the actors"*).

A privacy risk is *"the probability or "likelihood" and "consequence" of a loss of, or violation of, an individual's privacy"* (see Section 2.6.2). Thus, in the meta-model a privacy risk can be defined as the risk of re-identification occurring if data is released in its current format; the possibility of a breach of DPA and the impact of these.

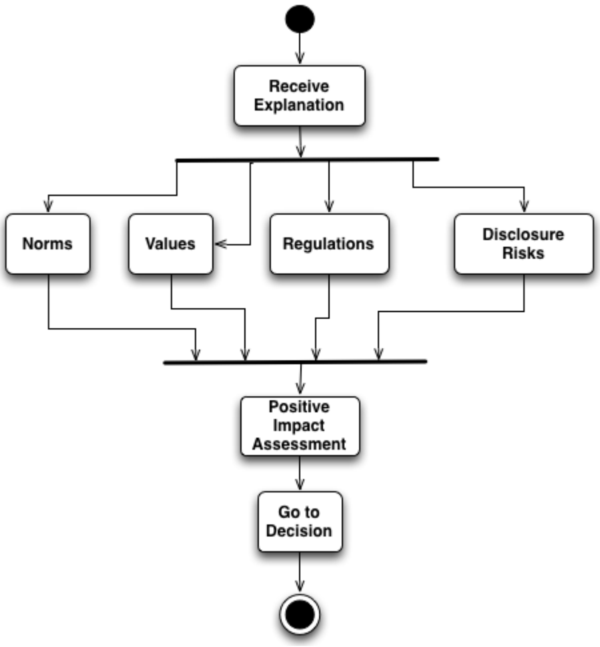


Figure 5.4: Activity Diagram : Risk Assessment

The risk assessment phase considers Nissenbaum's decision heuristics numbers five to eight, as these all relate to the evaluation or risk assessment of how the proposed new information flow will affect the privacy of the actors. Figure 5.4 shows the information gathered in the explanation phase from the; data, actors, transmission principles and prevailing context feeding in to the risk assessment and Figure 5.5 depicts the relations between these elements.

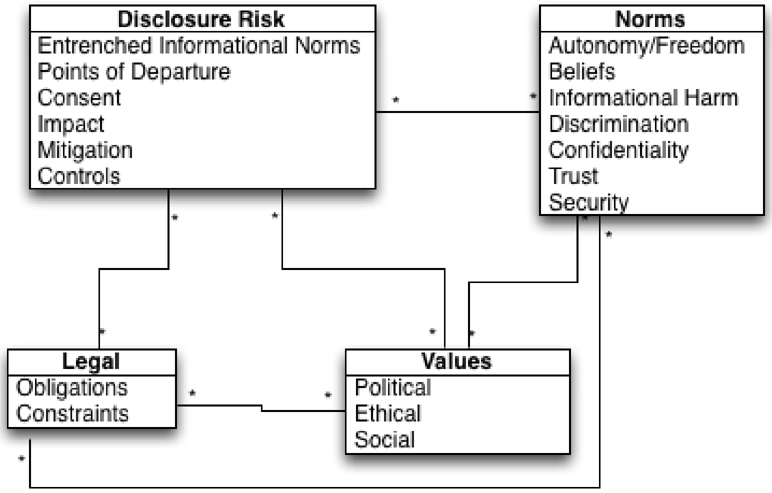


Figure 5.5: Risk Assessment - Class Relationships

The risk assessment asks for the following aspects to be considered:

Disclosure risk This looks at the information from the explanation phase and asking what the disclosure risk will be in light of:

- The entrenched informational norms;
- Any points departure from the entrenched informational norms identified;
- Whether consent has been granted and, if so, to what extent consent has been granted;
- The impact of any potential breach;
- Any proposed mitigation strategies put in place;
- Any proposed controls in place or to be put in place.

Norms These can be described as the standards or rules by which we conduct our lives. *”Norms define the duties, obligations, prerogatives, and privileges associated with particular roles, as well as acceptable and unacceptable behaviors”* (p. 133). They include invisible rules that most people abide by, such as not going out in public with no clothes on or not speaking out of turn. That could for example, involve not disclosing information given to us by a friend even if not explicitly told we can’t. Norms also include more formal rules, such as not discriminating against others. In terms of CI, the norms to be considered are the *”informational norms”* i.e. the norms that the actors will be expected to abide by in the capacity of their role. For example, a teacher may divulge a student's performance record to the student or their parents but would not be expected to divulge the same to other parents within the school. Thus, the informational norms to need to be considered in relation to the evaluation criterion to determine whether they may be infringed upon with the proposed new data flow. These are:

Autonomy/Freedom Referring to the freedom or autonomy of any of the actors, but perhaps most pertinent to the data subject as it is their information that may be released;

Beliefs This refers to any belief system, this may be religious, political or strong opinions held by any of the actors that could be affected by the proposed new data flow;

Informational harm This refers to the risks of informational harm to the data subject that may arise as a result of the new data flow;

Discrimination Referring to whether the new data flow could result in any form of discrimination;

Confidentiality Whether confidentiality may be breached as a result of the proposed new data flow;

Trust This asks whether publication could result in any breach of trust;

Security Referring to whether the new data flow could pose any security risks.

Legal This refers to any legal obligations that may be imposed on the public body in relation to the data. This may be constraints such as DPA regulations, or obligatory, such as the Re-use of Public Sector Information Regulations 2015 (ROPSIR) which require public bodies to publish data open source.

Values The values are *"the objects around which a context is oriented"*. These may be social, political or ethical values that could be affected or altered as a result of the proposed new information flow, asking whether it is possible that the proposed new data flow could impose an imbalance of some sort and thereby infringe on one or more of these values. These values will, of course, need to be considered in light of the norms and any legal or regulatory constraints or obligations. Thus, the relationship between each subclass within the risk assessment phase will always be many to many. This is because within each subclass there are multiple considerations, each of which will need to be considered in relation to all of the considerations within the other subclasses (see Figure 5.5).

The relationship between each subclass within the risk assessment phase will always be many to many. This is because within each subclass there are multiple considerations, each of which will need to be considered in relation to all of the considerations within the other subclasses. These relationships are depicted in Figure 5.5.

Once all of these considerations have been taken into account, the final step of the risk assessments involves conducting a positive impact assessment. The reason practitioners are asked to look for any potential positive impact is to capture relevant changes that, while they may represent a perceived *"deviation from entrenched norms"* or values could actually have a positive rather than negative impact. Take for example, a new technology innovation developed such as when the smart phone was first introduced. At the time, many people, including the primary researcher, felt the use of a smart phone posed too much of an infringement on personal privacy as users would be permanently contactable and locatable. However, most users now accept this fact and being contactable all the time could be argued to now be an *"entrenched norm"*. Thus, this shows that even entrenched norms and values may be subject to change over time and it is therefore important that positive, as well as negative impact be assessed. Therefore, the final step entails assessing:

Positive Values This should detail any positive values that publication will bring such as improvement in transparency or commercial gain; and

Overriding Values Outlining any overriding reason why publication should go ahead, such as a legal obligation to publish the data.

Validation - Risk Assessment

Relating this to our hypothetical PB and the strategic themes (see Section 4.5, Figure 4.5), the risk assessment phase is where they will assess the privacy risks associated with making the dataset available as open data. This will involve reviewing the privacy aspect within the strategy theme (see Figure 4.5) as follows:

Risk Based on the information collected as part of the explanation phase, the context and the relationships between the data, the actors and how the data is shared, the PB practitioner can identify any applicable privacy risks. For example, if one of the data processors is a personal friend of one of the data subjects within the dataset, this should be noted as a privacy risk as part of the assessment.

Once the disclosure risks have been identified in light of legal constraints, norms and values, PB needs to determine whether there are any mitigating steps that can be taken to make the data available in open format. Then, they can make an assessment of any associated privacy risks in publication and use this to assist them in making an informed decision as to whether a particular dataset can be published. This is captured in phase three, the decision.

5.4.3 Phase 3 - Decision

The third key element, *Prescription*, relates to the findings on whether a practice violates privacy. Prescription asks that the assessor to establish, based on the evaluation carried out in the Risk Assessment phase, whether or not privacy has been or is likely to be violated if the new process is implemented. This involves presenting the findings which will guide the practitioner in whether or not a practice or process poses a potential challenge to privacy. This, it is contended, involves making a decision as to the compatibility or non-compatibility of the information for allowing those changes or alterations in the data flow. Therefore, this has been translated in the meta-model into 'Decision'. The reason for this is that practitioners may not relate to what prescription means, whereas they will be familiar with decisions and making decisions (see Section 2.5).

Figure 5.1 shows an activity diagram depicting each phase of the meta-model and the order in which practitioners will progress through the framework (i.e. the activity flow), starting with phase 1, Explanation; moving on to phase 2, Risk Assessment; and culminating at phase 3, Decision.

With reference to the decision heuristics (DH), the final heuristic Nissenbaum asks readers to consider is whether, based on the findings made in the previous considerations; *"contextual integrity recommends for or against the proposed new practices"*. Again, this requires practitioners to make a decision in light of findings and therefore, this DH is considered in the final phase, decision.

In the decision phase there are only two classes. These classes and their relationships are shown in Figure 5.6.

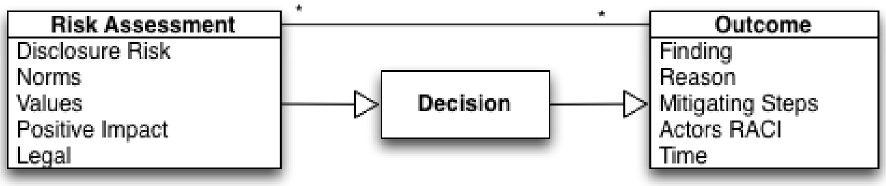


Figure 5.6: Decision - Classes

The attributes within each class will be:

Risk Assessment The Risk Assessment class will be carried over from Phase 2. This will feed into the decision element and produce an outcome class. This outcome class will hold the decision from each attribute group, contributing a score to aid the practitioner in making a decision. The relationship between the two classes is many to many as both classes contain multiple considerations that will need assessing.

Outcome The outcome class will consist of:

Finding This will be a ’to publish’ or ’not to publish’ decision.

Reason Here a series of contributing reasons can be recorded, this could, for example, refer to legal compliance such as ’Data Protection’ or a ’no privacy issues found’ reasoning.

If the decision is to publish the following additional attributes will also need to be completed:

Mitigating Steps This category will detail any mitigating steps that need to be carried out before publication can take place, this may, for instance, include redaction or anonymisation.

Actors Recording who is responsible, accountable, consulted and informed (RACI) as part of the process helps achieve transparency and provide assurance that proper process is followed in making, implementing and enforcing decisions made. Thus, here practitioners are asked to complete a responsibility matrix (Project Management Institute 2004). This will outline who is:

- Responsible; showing who is responsible for publication;
- Accountable; depicting who is accountable if the decision is challenged or there is a problem (there can only be one person accountable in a RACI matrix);
- Consulted; showing the actors whose opinions must be sought; and
- Informed; detailing which actors who need to be informed of the decision.

Time This will contain details of regularity of publication updates.

Validation - Decision

The decision phase for our hypothetical PB is where they will assess the privacy risks associated with making the dataset available as open data and identify any mitigation strategies. This will involve reviewing the privacy aspect within the strategy theme and any identify any ’how’s’ applicable in the actions theme (see Figure 4.5) as follows:

1. **Strategy:** *Mitigation* Here, the PB will identify what mitigation strategies there may be available for removing or reducing the identified risks, such as obfuscating attributes that contain identifiers (see 2.4.3) prior to publication;

2. **How:** *Data format* In this section the PB will record how they believe they can best implement the mitigation strategy. For the scenario of the friendship between the data processor and subject, this could for example, involve pseudonymisation or anonymisation of the dataset (see Section 2.4.2);

3. **How:** *Governance* Here the PB will record the outcome of the assessment to include the finding, the reasoning behind the decision, the mitigation strategies identified and whether or not these were applied and who will be responsible for what aspects of publication etc. (RACI) going forward.

5.5 Worked Example

This section seeks to provide an illustration of how the meta-model can be applied in practice by providing a worked example for each concept discussed above. This will take the form of a public body practitioner ('PP'). To give the PP some context we will give him the role of Data Officer and have him employed at the local Lending Library, applying the concepts to ascertain what privacy risks a hypothetical dataset, the 'Library Lending Register', will pose if it was to be published in open format.

5.5.1 Explanation

For the explanation phase (see 'Phase 1, Explanation" section in Figure 5.7), PP will record details of the Library Lending Register (i.e. the 'Open Dataset' in Figure 5.7).

Following the meta-model, this means PP will need to record details of the attribute category group (or column) that are contained within the dataset being assessed. For example, if PP is assessing the Lending Register of books on loan, PP will need to note details of each attribute category group (column within the database). This could, for example, include Customer ID, name, address and no of books on loan (see Table 5.1).

Customer ID	Name	Address	Books on loan
12345	Alice Smith	A Street	5
23456	Bob Jones	B Street	1
34567	Eve Evans	E Street	2

Table 5.1: Library Lending Register - extract

In addition to capturing this information, PP will also categorise each attribute type, so that each attribute type can be assessed for potential privacy risks, i.e. for attribute, what category group does that attribute belong to (personal identifier, quasi identifier, sensitive or non-sensitive, see (see 'Open Dataset' in Figure 5.7). In this example, the name will be a direct identifier; the customer ID and address will be quasi-identifiers (they can link back to the personal information if linked) and the number of books will be non-sensitive. Thus, most of these columns contain potential identifying information.

PP will also need to record information pertaining to who has been involved in processing the data and what their role(s) are (i.e. the 'Actors' and 'Roles' in Figure 5.7). This will involve looking at where the data originated from (i.e. which department), who works

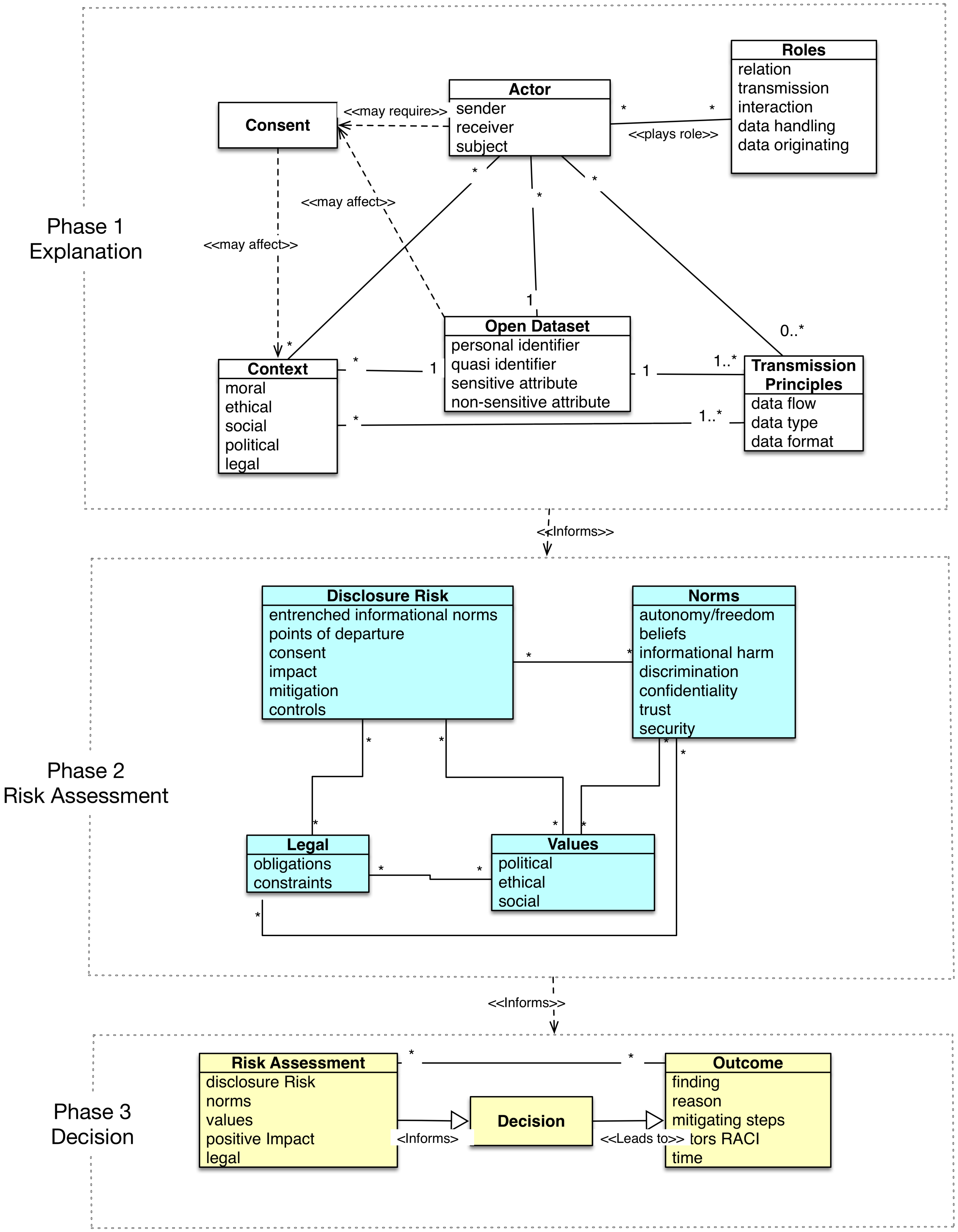


Figure 5.7: Meta-Model Overview

there and particularly, who worked with the data within that department (the data senders and data receivers). For the Library this might include a Learning Technologist and a Librarian. In addition, PP must ascertain who the responsible data controller is for the dataset. For context, PP must capture details of how these actors relate to each other and the data subjects (i.e. Alice, Bob and Eve, the library customers), e.g. are they personally acquainted or perhaps related? etc. Then, PP must record how the data is transmitted, what format it is held in, e.g. is it in a bespoke library lending database, a spreadsheet etc. Finally, PP will need to record how the data is transmitted, i.e. how does the information flow, both internally within the library and externally (the 'Transmission Principles' in Figure 5.7).

Once PP has recorded details of the data, actors, roles, the prevailing context, and the transmission principles, risks can be identified in light of established norms and values and any legal obligations.

5.5.2 Risk Assessment

The risk assessment phase is where PP will assess the privacy risks associated with making the dataset available as open data (see 'Phase 2, Risk Assessment' in Figure 5.7). This will involve reviewing the information captured as part of the evaluation and identifying any risks there might be associated with the data (see 'Disclosure Risk' in Figure 5.7). For example, if the Librarian and Eve are friends, PP would need to note this relationship as a potential risk as part of completing the risk assessment. For each of the risk assessment areas PP needs to note any associated privacy risks. The fact that these two actors are friends could have potential risk implications in a number of areas, meaning PP will need to consider the risks associated with each area:

Disclosure Risk : There could be a number of potential disclosure risks identified as a result of the friendship between the actors in this instance. For example, this could include the relationship between the actors giving rise to a consideration about who the librarian may share the data with and how appropriate such sharing may be. It should also consider what the repercussions would be if the librarian was to divulge information obtained in the course of their work to a third party such as another friend. Similarly, consideration will need to be given as to whether or not Eve has given consent to the data being processed and what that consent covers. Another risk associated with the friendship could be that the Librarian may divulge personal information about Bob (another library user) to Eve;

Norms : The friendship could result in Eve receiving preferential treatment such as being allowed extra books on loan (discrimination risk) or be privy to confidential information about Bob (trust and confidentiality risks);

Regulations : Adherence to data protection regulations. For instance, if Bob has not given consent for his data to be shared, the divulging of the information to Eve will constitute a breach of data protection regulations;

Values : A breach of confidentiality would infringe on social and ethical norms (see 'Risk Assessment' in Figure 5.7).

Once all of the disclosure risks have been identified in light of legal constraints, norms and values, PP will need to identify any mitigating steps that could be applied to facilitate making the data available in open format. Then, PP can assess whether, once these mitigations have been applied, any residual privacy risks could arise from publication and use this to inform the decision as to whether the Library Lending Register can be published.

5.5.3 Decision

The decision phase for our Lending Library is where PP must assess the privacy risks associated with making the dataset available as open data (see 'Phase 3, Decision' section in Figure 5.7) and, based on this, make an informed decision ('Decision' in Figure 5.7). As part of recording the 'Outcome' in Figure 5.7, PP will record the decision and the outcome of the assessment. This should outline what the finding from the assessment is; the reasoning behind the decision and the mitigation steps identified and whether or not these were applied and who will be responsible for what aspects of publication etc. (RACI) going forward. For the Privacy Risk Assessment carried out on the Library Lending Register, the mitigating steps could, for example, include:

- *Anonymisation* PP could advise that identifying attributes should be anonymised prior to publication Samarati (2001), Lablans et al. (2015);

- *Redaction* PP could recommend that personal identifiers such as names, be redacted prior to publication ICO (2012), Pfitzmann and Hansen (2010).

Once these steps have been completed PP will have a detailed record of the outcome of the privacy assessment that includes the finding and the reason for the decision. This will enable other practitioners within the lending library to refer to the decision made and provide the organisation with quality assurance and an audit trail of decisions made.

5.6 Meta-model - Conclusion

The meta-model created in this chapter, shows that it is possible to break CI down into its component parts (see Figure 5.7). In doing so, the meta-model shows how, by breaking CI down into logical phases and modelling how these interlink, a decision-flow can be established which can be followed in a methodical and systematic format when making decisions about privacy risks. The applicability of this meta-model in practice was demonstrated through a worked example that applied the concepts to a hypothetical public body, a public library (Section 5.5). Thus, the first proposition that *there are existing framework(s) that singularly or through amalgamation of concepts can be adapted to provide a practical foundation for determining privacy risks* hypothesis P1 (see Figure 3.2), holds true. The meta-model provides a practical example of how the CI conceptual framework can be translated into a working model for applying CI in practice.

5.6.1 Next Steps

The next step will be to determine how this model can be used to inform a working framework for applying CI in practice and answer RQ1 (see Figure 3.1 or Section 1.2). To this end the next Chapter explores how the meta-model can be used to create a more detailed framework that elaborates on each phase and breaks each phase down into a series of practical questions, based on the decision heuristics. These questions can then be used to apply CI in practice.

Chapter 6

Case Study - Contextual Integrity in practice

6.1 Introduction

It is argued in Chapter 1, Section 1.2 that existing privacy frameworks do not provide sufficient guidance on how to implement or conduct a privacy assessment, particularly in the context of open data publishing. Further, while practitioners are asked to consider context in some of these frameworks as part of the risk assessment (e.g. PIA, NIST, CI, see Chapter 2, Section 2.7), Chapter 4 confirmed there is little practical help in how to do so in practice.

In Chapter 5, proposition 1 (P1) was tested (see Figure 3.2) by taking an existing framework, Contextual Integrity (CI) and creating a visual model for how this could be adapted to work in practice. This resulted a meta-model based CI being created. This meta-model provided proof of concept that, in theory, CI can be adapted to provide a practical foundation for determining privacy risks (see Section 5.6). However, to demonstrate that there is little practical help available to practitioners requires more than proof of concept, it requires testing against what practitioners deal with in their daily work which is what this case study seeks to achieve.

This Chapter presents a case study on open data publishing. In the open data scenario, the recipient cannot be specifically defined as this would be anyone who downloads the data, something which none of the previous studies have considered (see Chapter 2, Section 2.7.7). Further, none of these studies have applied CI in a practical setting using real practitioners. This case study seeks to address this by trialling CI within a real practice setting, with existing circumstances where there is no clear demarcation between the context and the phenomenon being studied (Yin 2013) and is thus devised to apply the meta-model to a real problem in order to answer RQ1 (see Figure 3.1).

The intention is to take the meta-model and expand on this model so that this can be used to inform the creation of a privacy-specific decision framework that can support practitioners in making informed decisions about privacy before data is published as open data. The aim of this will be to guide practitioners through the CI framework in a step by step manner to ensure all aspects are considered within context, so that they can make an informed decision, and formulate a well-reasoned appraisal or refusal as to how and why that decision was arrived at.

The research approach for this will be a case study (see Chapter 3, Section 3.4.5) that utilises the meta-model to create a practical privacy-specific decision framework for applying the CI framework to a real problem, the publishing of public body information in open format.

The rest of this chapter is organised as follows. First, the case study protocol, detailing the method used for creating a paper prototype (see Chapter 3, Section 3.4.11) of the meta-model is presented in Section 6.2. This is followed in Section 6.3 by a description for how the meta-model was further interpreted to create a practical questionnaire for assessing the privacy risks of publishing open data, CLIFOD. In Section 6.4, the CLIFOD questions are evaluated by a group of peers, followed by applying CLIFOD in practice in a practical trial in Section 6.5. In Section 6.6, the findings from this case study are presented and the chapter concludes in Section 6.7, confirming how the resulting questionnaire provides the answer to RQ1.

6.2 CLIFOD Protocol

This case study used the meta-model as the starting point to create a working framework that applies ContextuaL Integrity For Open Data in practice (CLIFOD). CLIFOD is a practical paper prototype, in the form of a 98 question questionnaire, based on CI, designed to determine whether this prototype can provide an effective tool for practitioners in deciding whether or not a particular dataset is suitable for open data publication.

The research method chosen for this work is case study, in line with the overall methodology (see Chapter 3, Section 3.4.5).

The aim of conducting this case study is to establish whether or not Nissenbaum's Contextual Integrity (CI) framework can be successfully applied as a decision tool around privacy in practice when publishing public data open source. The intention is to establish whether the CI framework is appropriate for use as it is when applied in practice or, whether it will be necessary to adopt this to better suit the requirements of open source publication. This will be tested by applying the meta-model interpretation of the CI framework to real data in a practical setting.

6.2.1 Problem Definition

This study seeks to apply the meta-model in practice by creating a privacy-specific decision making framework that practitioners can adopt to support privacy decision making in practice (ContextuaL Integrity For Open Data in practice (CLIFOD)). This needs to be a framework that can be applied by individual Information Officers and Practitioners to guide them in their privacy decision making process.

It is contended that by using this framework, practitioners will be able to review privacy in context before publication takes place, enabling them to make an informed decision as to whether a dataset has any privacy implications. This in turn will allow practitioners to formulate a reasoned argument for why that particular dataset should, or should not, be published as open data.

6.2.2 Research Question and Proposition

The question this case study seeks to answer is RQ1 (see Chapter 1, Section 1.2) to demonstrate how an existing privacy framework, Contextual Integrity, can be adapted to become a useful prototype tool for supporting privacy decision making in practice. The CLIFOD framework will be created for the *public open sector* domain (see Section 1.1.1), the sector chosen to test the effectiveness of the adaptation of CI into a privacy decision making support tool. The intention is that future work will look to expand on this to make the final contribution of this thesis equally applicable and useful across multiple industry sectors, i.e. domain neutral.

6.2.3 Unit of Analysis

For the practical trial or application of CLIFOD, the unit of analysis will be *the privacy practitioner(s)* (see Section 3.5.2) within a public body. A public body is a body or organisation that is governed by public law which exercises public functions that are woven ”into the fabric of public regulation” (Burton 2002). Thus, this may be a central government body, a local government body or, indeed a university or other public organisation. For this case study, it seemed appropriate to find a public body that already publish some data in open format as they will have an existing process in place against which the CLIFOD can be tested.

In earlier research, local authorities were interviewed to establish their current practices for publishing data sources (Henriksen-Bulmer 2016). Thus, the privacy practitioners at a UK Local Authority, who currently publish date in open format, were approached to establish whether they would agree to work apply the CLIFOD framework to real data, in a real practical setting. They kindly agreed to collaborate and trial the CLIFOD framework.

Two practitioners who work with open data publication within the LA took part in the trial, one had a technical background and one whose background is process and policy. Both practitioners however, are actively involved in open data publication and making decisions about the suitability of datasets for open data publication. The practitioners confirmed that no formal process was in place within the LA for assessing privacy risks of data prior to publication. The current process was that the open data practitioner would assess privacy by conducting some basic checking to ensure no obvious breach of data protection. However, ultimate responsibility for data protection remained with the data owners (i.e. the originating department).

6.2.4 Data Collection Approach

Currently, the CI framework consists of a number of elements to consider and related discussion points. Nissenbaum discusses CI in terms of the questions that one might ask in applying the framework and discusses at length how this might be applied to particular situations. She does not however, provide much detailed practical guidance that users can easily apply to their particular scenario. To address this, CLIFOD will consist of a step-by-step approach with specific questions that practitioners need to consider.

Therefore, the first step will be to create the questions that practitioners will be asked to answer. This will involve expanding on the meta-model to create a more detailed framework or questionnaire that can be followed in a systematic manner. Thus, for the purposes of this case study, the meta-model and Nissenbaum's framework will be interpreted and broken down further into ContextuaL Integrity For Open Data in practice (CLIFOD), a privacy-specific questionnaire based on the meta-model and CI.This will take the form of a questionnaire that goes more in-depth in interpreting the meta-model and CI to elicit what information needs capturing, thereby enabling enable specific questions to be devised that align with both the meta-model and CI. Further, this will also involve creating an methodology for following CLIFOD, based on the the meta-model and CI frameworks. This will be done in three steps:

Step 1 Based on the meta-model and CI, define the design and layout of CLIFOD and devise the questions for applying the CI framework in practice.This step involved explaining how the meta-model has been interpreted and applied into CLIFOD to create a step-by-step questionnaire based on the meta-model and CI (see Section 6.3);

Step 2 Evaluate the model using 3 independent reviewers, a solicitor, a practitioner and an academic (*the peer group*, see Section 6.2.5); and

Step 3 Practical trial applying CLIFOD to real data, collaborating with a UK public body, a Local Authority (see Section 6.2.5).

6.2.5 Case Study Validation

Yin suggests that the best form of validation is to triangulate, i.e. seek review from three independent reviewers, each from a different perspective (Yin 2013). Thus, to validate the questions and ensure these are appropriate, non-biased and in line with Nissenbaum's CI framework and the meta-model, the CLIFOD questionnaire will be first be reviewed and validated by three independent parties, each of whom will review CLIFOD from a different perspective (*the peer group*) .

Step 2 - Peer Group Validation

Once the questions have been devised these will be sent to three independent reviewers for comment and review. The choice of reviewer had been made to achieve triangulation in that, each reviewer was asked to review the questions from a different perspective:

Reviewer 1 was asked to review the questions from a legal perspective (a solicitor);

Reviewer 2 was asked to comment from a practical standpoint (a practitioner);

Reviewer 3 was asked to comment on the questions in general and to review from a methodological perspective (an academic).

Each reviewer will be selected for their expertise within their field (see also Section 6.2.4). A copy of the emails sent to the reviewers can be found in Appendix B, Section B.2).

Solicitor

The legal aspect of the framework was reviewed and evaluated by a solicitor who is currently working within a public body with experience of freedom of information and data requests from the public. In addition to commenting on the questions themselves and their suitability, the solicitor was also asked to identify any areas that, if not addressed or included (if omitted), could potentially leave a public body open to challenge if publication was refused.

Practitioner

The practical aspect of the framework was reviewed and evaluated by a practitioner who works with, and makes decisions around, public sector information on a daily basis. To ensure there was no bias from the local authority participating in the case study, an Information Governance Officer from a different local authority was approached to evaluate the questions. This practitioner was asked to review the questions for suitability and, in addition, to consider these from a practical perspective were his local authority to apply them in practice and comment on their practical relevance and applicability.

Academic

For general comment an Academic was be asked to review and evaluate the questions from a suitability perspective and from a methodological perspective. The academic approached is an experienced researcher with a specialism in methodology and the research process.

With regards to suitability of the questions, each of the reviewers will be asked to highlight and identify:

1. Any areas/questions that had been omitted;
2. Identify areas that required further explanation or elaboration;
3. Check that questions followed logical progression and where in appropriate sections;
4. Any areas that may be difficult to comprehend; or
5. Any other comments or observations.

Step 3 - Practical Trial Validation

For Step 3, the practical trial, the intention is that CLIFOD will be trialled and used to assist practitioners in the decision making process when considering whether or not to publish a dataset as Open Data. To this end, a practical trial of CLIFOD will be carried out, using real data in a real setting. During this trial, data will be collected through think aloud and contextual interviews.

Prior to the practical trial, the practitioners who had agreed to collaborate in the Case Study will be provided with some background information (see Appendix B, Section B.1) and asked to sign a participant agreement form (see Appendix A, Section A.3). They will also be asked to prepare three different types of datasets for CLIFOD assessment as follows:

1. **Dataset 1** this should be a dataset that has already been published;
2. **Dataset 2** this should be a dataset currently under consideration for potential open data publication;
3. **Dataset 3** this should be a dataset proposed for open data publication but which the practitioner believes is not likely to be suitable.

The reason for selecting these three types of dataset is to ensure CLIFOD is trialled on different types of data with varying potential privacy implications and risks.

During the trial, the practitioners will be supplied with a spread sheet containing the framework questions and asked to apply these to real data under consideration for open data publication. Because these interviews cannot be conducted face-to-face, due to distance and time constraints, a second, complimentary method of data collection will also be used, think aloud. Thus, the interviews will be conducted through a series conference calls via telephone and Skype with the local authority. As the practitioners apply the framework, they will be asked to explain out loud their thinking and reasoning. The interviewer (the primary researcher) will listen and observe, making notes of the thought and decision processes that the practitioner applies.

Think aloud, involves users talking out loud while completing a task, explaining what they are doing and why. This allows researchers to capture the thought process of the user as they work (Ericsson and Simon 1980, Davison et al. 1997). Contextual interviews involve observing participants as they work (Beyer and Holtzblatt 1998). For further detail on these methods, see Section 3.5.2.

Three datasets will be evaluated by running them through the CLIFOD questionnaire (the framework). This will be done as a collaboration exercise between the privacy practitioners and the researcher. The intention is that three datasets will be assessed using CLIFOD for the practical trial. These will consist of three types of data as follows:

1. One dataset that has already been published;
2. A dataset not published as yet but being considered as a candidate dataset for open data publication;
3. A dataset that is unlikely to be considered suitable for open data publication in the opinion of the practitioner but which nevertheless, will require assessment.

The reason three different types of datasets have been suggested is to ensure a variety of different datatypes are considered using CLIFOD. This will ensure differing levels of privacy risk is considered as part of the trial and thus, validate or not whether CLIFOD can be used as an effective tool in assessing privacy risks in both straightforward and complex cases.

For validation of the findings of the trial process and method of data collection, the answers noted by the practitioners on their spread sheet will be compared with the answers noted by the interviewer during the observation itself.

6.2.6 Analysis

The outcome of the trial will be interpreted and analysed to establish to what extent privacy has been preserved and/or compromised within the datasets considered.

6.2.7 Manner of presentation

The findings will be presented in this report.

6.3 Step 1 - Creating the CLIFOD Questionnaire

The meta-model identifies two areas where Nissenbaum offers some level of detailed guidance for applying the framework to a practical scenario in the book (Nissenbaum 2010). These are:

Key Elements Nissenbaum contends the CI framework consists of 3 key elements; Explanation, evaluation, and prescription" (p. 190); and

Decision Heuristics the; "Augmented Contextual Integrity Decision Heuristics" consisting of nine steps (see Chapter 5, Section 5.3.5).

Whilst these two areas do provide more tangible advice on the application of the framework, they still do not go into sufficient detail for practitioners to apply without considerable interpretation and breaking this down into more tangible individual elements.

In the meta-model, the first step involved translating the key elements into three phases: explanation, risk assessment and decision (see Figure 5.7). The decision heuristics were then allocated according to which phase they fell into.

However, for CLIFOD, a more detailed interpretation of each of the decision heuristics (DH) will also be necessary to create the desired systematic framework or questionnaire. This deeper analysis of CI will change the meta-model from a model to an actual questionnaire that will enable practitioners to work their way through the framework in a more gradational manner. These questions and the reasoning behind them are detailed in Section 6.3.3.

6.3.1 Decision Heuristics

Nissenbaum provides nine decision heuristics as an approach to evaluating a system or practice. This involves assessing the information in view of pertinent "values, ends and purposes". Nissenbaum contends that comparing existing flows of information with the proposed new flows, accounting for any breaches (or potential breaches) of values and comparing these to any potential privacy conflict or threat, will afford practitioners the opportunity to identify and mitigate against these. This will in turn enable practitioners to "establish the significance of each value in light of its contextual ends and purposes" (p 191).

Once the overarching areas had been developed in the meta-model (see Figure 5.1), the next step is to translate the groups of DH allocated to each phase and devise suitable questions to capture the knowledge or information required within each of these three phases based on their allocated DH.

6.3.2 Definition/interpretation

Before explaining the logic behind each group of however, some definitions of some of the terminology used by Nissenbaum when she discussed information flows might prove helpful. Information flows are discussed by Nissenbaum in terms of; ”actors and roles” (the people involved with the data and their role); ”the attributes” (i.e. the information itself); and, how this information is transmitted (”transmission principles”). Some of these terms may not be familiar to practitioners and therefore, where this is the case, terminology that will use in practice has been used in CLIFOD. The reasoning behind each term and why that particular choice of terminology was chosen is provided.

Actors

Nissenbaum refers to the people involved with the data as ”Actors”. There are three categories of actors; (1) ”information subjects”; (2) ”data senders”; or (3) ”data recipients”. In practice, this will be the people who handle the data or about whom the data pertains. Therefore, it seemed more appropriate to relate this to the people who handle and data in a public body. The public body will likely relate to the definitions laid down in DPA and therefore, these will be referred to in CLIFOD in accordance with established convention under The Data Protection Act 1998 (DPA);

1. The information subject(s), i.e. the person(s) the data collected pertains to. DPA uses the ”data subject” when referring to the ’information subject’. Thus, this is the term that will be utilised under CLIFOD;

2. The data sender(s); here there are three Actors that are accountable, responsible and/or handle the data:

   (a) The ’public body’ (organisation). The reason for including the public body itself is that, in legal terms, the organisation is a legal entity (person) with rights and obligations (Friedmann 1951), and therefore, needs to be considered as, ultimately, the organisation is accountable if any decision to publish is challenged;

   (b) The ”data controllers”, i.e. the people who make decisions about the data; ’the data controllers’, those responsible for making decisions around what processing may or may not be permitted (DPA 1998, s. 1(1)); and

(c) The ”data processors”, i.e. the people who process or handle the data on a day to day basis in accordance with the directions given by the data controller (DPA 1998, s. 1(1));

3. The ”data recipients”, i.e. the end users, those that download, manipulate and utilise the data once it has been published.

Information/Data

Nissenbaum refers to personal information in terms of ”attributes” (p. 166). This is in line with database management terminology and will be understood by practitioners as they will store, control and manipulate the information that will be under consideration when applying CLIFOD in databases, these databases will vary depending on the public body and the individual department but nonetheless, the information (data) will be held in a database of sorts. However, different areas will use different terms, e.g. in information security they refer to the elements of a system as assets (Faily and Fléchais 2010).

Therefore, when referring to the data itself, it makes sense to use the terminology adopted in DPA and database management theory as this terminology will be familiar to practitioners.

6.3.3 CLIFOD questions

Starting with how the meta-model have divided Nissenbaum’s nine DH into the three phases (see Section 5.3), these have been broken down into more detailed, specific questions. This was done by taking the original DH questions, interpreting and elaborating on the meaning of this to devise specific questions that could be applied in a practical setting. Each question and the thought process for the resulting question is explained in turn.

In the meta-model, each phase has been broken into sub-sections based on the data to be captured and related DH allocated to each step (see Figure 5.7). To translate this into the more detailed questions required for CLIFOD, for each phase within the meta-model, the group of DH related will be dealt with in groups within their the meta-model phase below.

At the beginning of each phase, each DH question will be repeated, followed by an explanation for CLIFOD and the questions devised relating to that phase in the meta-model.

6.3.4 Meta-model Phase 1 - Explanation

The meta-model Explanation phase (Phase 1) encompassed four areas: open dataset, actors, context and transmission principles (see Figure 5.2). Beneath that, the meta-model included the first four DH's as relating to these categories (see Section 5.4.1 and Figure 5.3). These DH are discussed and elaborated upon in CLIFOD as follows:

Decision Heuristic 1

> ”Describe the new practice in terms of information flows” (Nissenbaum 2010).

In order to understand the information flows however, it is necessary to first understand the data. What is the data about and what are the individual parts that make up the datasets. Then, questions can be asked about the existing information flows, i.e. how the data is currently processed. As this relates to finding out about the data itself, this has been placed in the explanation section.

The following questions were devised to cover this area:

Describe dataset What is the dataset about?

Describe Attributes What attributes are in the data? - please provide a full list

For each attribute (set) within the dataset, please describe:

Q1 Do any of the attributes contain personal information (identifiers)?

Q2 Describe the attribute, exactly what data is collected?

Q3 How many fields/pockets within the dataset contain this type of information?

Q4 Do any of the attributes contain quasi-identifiable data (identifiable through inference data)?

Q5 Describe the attribute, exactly what data is collected?

Q6 How many fields/pockets within the dataset contain this type of information?

Q7 Do any of the attributes contain individual specific attributes (sensitive data)?

Q8 Describe the attribute, exactly what data is collected?

Q9 How many fields/pockets within the dataset contain this type of information?

Q10 Do any of the attributes contain other non-identifiable information (non-sensitive)?

Q11 Describe the attribute, exactly what data is collected?

Q12 How many fields/pockets within the dataset contain this type of information?

Decision Heuristic 2

> ”Identify the prevailing context. Establish context at a familiar level of generality (e.g., health care) and identify potential impacts from contexts nested within it, such as teaching hospital.” (Nissenbaum 2010).

This relates to the organisation, what type of organisation and in what context was the data collected. Again, these are details about the organisational context so this was placed in explanation section. The following questions were devised to cover this area:

Q13 How is data currently processed?

Q14 Where does data originate? (Department/external partner)

Q15 Is this data original to this dataset or was data collected from another department/processor?

Q16 If yes, who/which department does the data originate from?

Decision Heuristic 3

> ”Identify information subjects, senders, and recipients” (Nissenbaum 2010).

This relates to the actors, i.e. the people that handle or are the subjects of the data. Therefore, this was placed in the explanation section. The questions relating to these have been phrased as follows within CLIFOD:

Public Body

Q17 Who owns the information? (public body, third party)

Q18 If owed by a public body, what is the public body that owns the information?

Q19 What it the role of the public body/department? (e.g. library, local authority, University etc.)

Q20 What is the relationship between the public body and the data subject?

Data Subject

Q21 Who/what is the data subject(s)?

Q22 What is the role of the data subject in relation to the data? (e.g. borrower, employee, citizen)

Q23 If data subject(s) is a person what relationship does this person have to the data processor(s)? (none, co-worker, friend, family member, citizen, professional, client, etc.) If multiple, please state all that refer

Q24 If data subject(s) is a person what relationship does this person have to the data controller(s)? (none, co-worker, friend, family member, citizen, professional, client, etc.) If multiple, please state all that refer

Q25 In what context did the data subject divulge the information?

Q26 Was there a legal obligation on the data subject to divulging the information?

Q27 How does data subject interact with data processor/originator? (e.g. friend, co-worker, professionally, citizen, employment) If multiple, please state all that refer

Data Originator

The person who originally collected the information (sender). This would be a data controller but may be a different data controller to the one who is considering whether or not to publish the data, therefore, this has been included with a slightly different name to distinguish from any subsequent data controller(s) who may be making the decisions around the data):

Q28 Who collected the data? (originator) (if multiple, please answer questions for each originator/controller)

Q29 What is the role of the data originator in relation to the data? (librarian, information officer, employee, third-party partner employee etc.)

Data Controller

Q30 Is data controller also data originator?

Q31 Who is the data controller(s)? (if multiple, please answer questions for each controller)

Q32 What is the role of the data controller in relation to the data? (librarian, information officer, employee, third-party partner employee etc.)

Q33 How does data controller interact with data subject(s)? (e.g. friend, co-worker, professionally, citizen, employment)

Q34 How does data controller interact with data processor? (e.g. friend, co-worker, professionally, citizen, employment)

Q35 How does data controller interact with data recipient(s)? (e.g. friend, co-worker, professionally, citizen, employment)

Data Processor

Q36 How many data processors are/have handling/ed the data? For each processor please state:

Q37 What is the role of the data processor in relation to the data? (librarian, information officer, employee, third-party partner employee etc.)

Q38 In what capacity was the data collected by the data processor(s)?

Q39 What is the data processor(s) position in the organisation?

Q40 How does data processor interact with data subject(s)? (e.g. friend, co-worker, professionally, citizen, employment)

Q41 How does data processor interact with data controller? (e.g. friend, co-worker, professionally, citizen, employment)

Q42 How does data processor interact with data recipient(s)? (e.g. friend, co-worker, professionally, citizen, employment)

Data Recipient (user)

Q43 Who is the data recipient(s)?

Q44 What is the role of the data recipient(s) in relation to the data? (librarian, information officer, employee, third-party partner employee, unknown etc.)

Q45 What position does the recipient(s) hold?

Q46 How does data recipient interact with data subject(s)? (e.g. friend, co-worker, professionally, citizen, employment)

Q47 How does data recipient(s) interact with data processor(s)? (e.g. friend, co-worker, professionally, citizen, employment)

Decision Heuristic 4

”Identify transmission principles” (Nissenbaum 2010).

When Nissenbaum discussed how the data is conveyed between the Actors, she refers to this as the ”transmission principles”. Nissenbaum contends that merely considering the data and the actors is not sufficient, consideration also needs to be given to how the data is transmitted and the context in which that transmission takes place. This context needs to include what restrictions may be in place for the data flow and how this is managed and enforced. However, for the purposes of this study, the transmission principle will be open data publishing and therefore, a presumption that no restrictions on the recipients ability to re-use, process and manipulate the data once received will apply. What will be considered will be the format and type of data (dynamic or static) and how often it is proposed this is updated. Again, this relates to factual information about the data and how this is conveyed and therefore, this has been placed in the explanation section. The following questions will be asked in this category:

Q48 How will the data be shared/published?

Q49 In what format will the data be shared?

Q50 Will data published be static (i.e. a cut of the data at specific point in time) or dynamic (e.g. real time and changing regularly)?

Q51 How often will it be updated?

Q52 When updated, will the existing data be replaced or will version control or dates be used to denote updates?

6.3.5 Meta-model Phase 2 - Risk Assessment

The meta-model phase 2 encapsulated five areas: norms, values, regulations, disclosure risks and positive impact assessment (see Figure 5.4). Beneath that, the meta-model identified DH five to eight as relevant to these categories. These DH are discussed and elaborated upon below.

Decision Heuristic 5

> ”Locate applicable entrenched informational norms and identify significant points of departure” (Nissenbaum 2010).

Both the social and data etiquette will need to be considered as well as any legal and ethical considerations that need to be taken into account in order to establish the prevailing context. Arguably, this could relate to either explanation or risk assessment as it relates partly to information about the data and the purpose of collecting it, and, partly to the surrounding circumstances which will form part of the risk considerations that will need to be considered. However, a decision was taken that this category is still primarily concerned with gathering information about the data and thus, this was placed in the explanation section. The following questions were devised to capture this category:

Q53 What was the original purpose of the data collection/processing?

Q54 What was the social context of the data collection? e.g. school would be educational context, council tax would be for tax collection etc.

Q55 Was permission/consent sought for processing of the data from data subject(s)?

Q56 Was the data collected with a view to process beyond its original purpose?

Q57 If yes, what was that purpose?

Q58 Was data collected direct from the data subject(s)? (i.e. did they provide the information)

Q59 Was consent sought for original collection/processing purpose?

Q60 If yes, was consent granted for secondary processing purpose?

Q61 If yes, was consent granted for specific secondary processing purpose?

Q62 If yes, what was that purpose?

Q63 Were there any limitations on secondary purpose to which consent was given?

Q64 If yes, what were those limitations?

Q65 Was consent granted for open re-use/sharing/processing?

Q66 Are there any overriding considerations as to why secondary processing should be allowed despite lack of consent/limited consent?

Q67 If yes, what are those considerations?

Q68 Do any of these consideration have legal authority?

Decision Heuristic 6

> ”Prima facie assessment: There may be various ways a system or practice defines entrenched norms. One common source is a discrepancy in one or more of the key parameters. Another is that the existing normative structure for the context in question might be ”incomplete” in relation to the activities in question.... A breach of informational norms yields a prima facie judgment that contextual integrity has been violated because presumption favours the entrenched practice” (Nissenbaum 2010).

This was interpreted as an initial evaluation of the explanations gathered in that it asks for thought to be given as to whether those explanations in themselves have identified any potential risks or indeed, whether the context surrounding the collection of the data pose a privacy risk. Therefore, this will be captured through gathering and scoring each sub-category (data; information flows; actors; transmission principles and context) so as to obtain an overview of the dataset and its privacy implications. Further, for each sub-category, risks and mitigation strategies will be identified and listed. Finally, practitioners will be given an opportunity to amend the initial score if they decide to apply the identified mitigation and thus, reduce the risk.

Decision Heuristic 7

> ”Evaluation I: Consider moral and political factors affected by the practice in question. What might be the harms, the threats to autonomy and freedom? What might be the effects on power structures, implications for justice, fairness, equality, social hierarchy, democracy, and so on? In some instances the results may overwhelmingly favor either accepting or rejecting the system or practice under study; in most of the controversial cases an array of actors emerge requiring further consideration” (Nissenbaum 2010).

This, in the meta-model, falls into the risk assessment section. It has been placed here as practitioners are asked to start considering the decisions reached in question 6 further. This section refers to the broader context by asking practitioners to now also take into account individual rights, values and norms as well as the surrounding social, political and moral contexts. Within CLIFOD, the following questions will be asked to cover this DH:

Assess disclosure risk

Q69 Have any disclosure risks been identified? (please list all that apply)

Q70 Please detail and identify any disclosure control processes that may be relevant?

Roles

Q71 Is data subject(s) aware of data being published?

Q72 Has data subject(s) consented to disclosure?

Q73 Has data controller(s) consented to disclosure?

Attributes

Q74 What are the risks of the attributes, once published, being linked to external data in such a way that they can pose a risk of contributing to re-identification of a data subject? (please be specific as to what risk and how this might pose a new risk)

Values

Q75 How could publication be perceived to infringe on any political values? (please also explain whose values and how they might be infringed upon)

Q76 Would publication impose power imbalance and thus, infringe on any moral values? (please explain what those moral values are, whose moral values and how these might be infringed upon)

Q77 Could publication be perceived to impose any form of power imbalance? (please explain how such a power balance might arise, who might be affected and how these might be imposed)

Q78 Could publication be perceived to infringe upon any social values? (please explain what those values are, whose values they are and how these might be infringed upon)

Norms

Q79 Would publication pose a threat to the autonomy or freedom of the data subject(s)? (please explain how this might pose a threat, who could be affected and how)

Q80 Are there any belief systems that may adversely be affected by publication? (please explain what belief system could be affected, how and who would might be adversely affected)

Q81 Would publication result in any form of discrimination (please explain what form of discrimination, how this might occur and who would might be adversely affected)

Q82 Could publication result in informational harm on the data subject(s)? (if yes, please clarify how such harm could occur, who would be affected and how)

Q83 Is there any risk that publication could result in breach of confidentiality? (if yes, please clarify how such a breach might occur, who would be affected and how)

Q84 Could publication result in breach of trust (if yes, please clarify how such harm could occur, who would be affected and how)

Q85 Would publication infringe on any legal compliance? (if yes, please clarify what legal compliance might be breached, how and who would be affected)

Q86 Would publication impose any security risks (if yes, please clarify what the security risk is, how and who would be affected)

Decision Heuristic 8

> ”Evaluation II: Ask how the system or practices directly impinge on values, goals, and ends of the context. In addition, consider the meaning or significance of moral and political factors in light of contextual values, ends, purposes, and goals. In other words, what do harms, or threats to autonomy and freedom, or perturbations in power structures and justice mean in relation to this context?” (Nissenbaum 2010).

Here Nissenbaum asks that consideration be given to what potential consequences there might be if any of the Risks, roles or values identified in Question 7 were to occur. This may be either negative or positive. Thus, this forms part of the risk assessment and, as such, has been placed in phase 2, Risk Assessment. The questions here follow on and elaborate on the context and potential harm identified through the questions asked in the previous sections:

Q87 Is there a reasonable expectation on the part of the data subject and/or data controller of data being kept confidential and not published?

Q88 Are there any privileges or prerogatives that arise from processing or publishing the data from which the public body, the data processor, controller or originator may benefit or be seen to benefit as a result of publication?

Q89 What are the positive values that publication will bring/enhance? These may include commercial gain, improved transparency, meeting legal obligation etc.

Q90 Are there any overriding legal, moral or ethical reasons why publication should be allowed

6.3.6 Meta-model Phase 3 - Decision

The meta-model phase 3 seeks to capture the decision to be made based on the findings of the first two phases (see Figure 5.6). This consist of only one category, the ’decision’ itself and this, in turn, seeks to capture the last DH, number 9. This is discussed and elaborated upon in CLIFOD as follows:

Decision Heuristic 9

> ”On the basis of these findings, contextual integrity recommends in favour of or against systems or practices under study. (In rare circumstances, there might be cases that a re-sustained in spite of these findings, accepting resulting threats to the continuing existence of the context itself as a viable social unit” (Nissenbaum 2010).

The final question in Nissenbaum’s nine steps relate to making a decision, thus, this has been placed in phase 3, Decision.

In CLIFOD, this decision will be to publish or not to publish. Therefore, the questions have been designed to explain the decision taken, what further actions may need to be carried out prior to publication, who is responsible and who will be accountable. Further, if the decision is not to publish, CLIFOD asks practitioners to explain the decision so that this may be used to defend or explain the reasoning, in case the information is requested or the decision challenged in future:

Decision In light of the findings what decision has been reached:

If decision is to publish

Q91 What work will need to be undertaken prior to publication if any?

Q92 If decision is to publish part of the data rather than the raw data, please specify reason for this decision?

Q93 What is the timescale for publication?

Q94 What processes have/will be put in place for updating of the information?

Q95 Who will be responsible for publication?

Q96 Who will need to be notified of publication?

Q97 Who will be answerable if any questions or challenges arise as a result of publication?

If decision is not to publish

Q98 What is the reason for non-publication? please provide a full statement including any legal reference where legal constraints form part of the reasoning. This statement should be copiable so that this may be used by data controllers and processors in defence of refusing any request for the release of the data

By breaking these nine steps into bite-size specific individual questions it is hoped, this will make CLIFOD more useful in practice by guiding practitioners through the CI framework in a step-by-step manner.

6.3.7 Unit of analysis - Scoring

As part of the CLIFOD assessment and to guide practitioners on progress, some form of rating or scoring system will also be needed to allow practitioners to assess how one value compares with another. As CI is a subjective assessment it is not intended that any scoring mechanism will calculate or objectively make any of the decisions around whether a dataset is suitable for open source publishing. Rather, scoring is intended as an progressive visual aid to assist practitioners in making the decisions. Thus, a unit of analysis for the CLIFOD questions will need to be devised.

There are many ways risk can be scored depending on the subject to be assessed and the risks identified, indeed there are standards for how to define risks (Cagliano et al. 2015).

In this instance, defining the values of the scoring is no simple task as all the questions are subjective, very few are objective. Therefore, the scores for each question cannot be pre-set to a score as the answer will depend, not only on the data, but also how risk averse a particular public body or indeed data controller may be.

In practice, risk assessment involves; risk identification; risk analysis and risk evaluation (ANSI 2011) with each risk identified scored on the basis of likelihood and severity or impact or risk level and action. This can then be depicted using colour coding (e.g. a traffic light system) and/or a scoring matrix (Lyon and Popov 2016b, Heiser 2008), as used in project management (Hall 2011).

What we are trying to establish in the case study is; (a) whether the privacy of the data subject(s) might be violated; and (b) the potential impact this might have. Therefore, for the purposes of this case study, we will adopt the ratings of; high (red); medium (amber); or low (green) for both violation and impact. The scoring of the risk will take place as part of the risk assessment, with scores colour coded (red, amber, green) for easy overview.

Practitioners will then be asked to make the assessment and score each risk identified. In addition, a section will be provided for mitigation strategies. Here practitioners will be encouraged to note any mitigation strategies that may be relevant and re-assess the risk if those strategies are adopted. This will allow the practitioners to consider whether for example, the data can be anonymised sufficiently that publication can still take place. These scores will assist the practitioner with deciding whether the case for open data publication can be met.

6.3.8 Format and layout of CLIFOD

CLIFOD will be presented to the practitioner in a spreadsheet consisting of three worksheets, one for each phase.

At the header of each worksheet will be the name of the dataset under consideration. The layout of the worksheets will be as follows:

Tab 1 - Explanation

This tab will consist of 5 columns with the questions organised in groups according to the subject header under consideration in the rows below. The column headings are:

1. Step - this will be broken into numbers corresponding with the group of questions.
2. Contextual integrity - this will hold the subject category for each group of questions and a brief explanation of what that section relates to e.g. ”describe the existing practice in terms of information flows”
3. Applying CI framework to dataset - this will contain a brief overview of the subject, e.g. data, information flow, actors etc.
4. Questions - here the questions will be listed
5. Answers - here the practitioner will be able to respond to the questions

Tab 2 - Risk Assessment

This tab will consist of 8 columns with the same structured rows below. The columns will follow a similar pattern to that outlined for tab 1 as follows:

1. Step - this will be broken into numbers corresponding with the group of questions;
2. Applying CI framework to dataset - this will contain a brief overview of the corresponding groupings from tab 1, the information having been carried over to avoid the practitioner needing to repeat themselves e.g. data, information flow, actors etc.;
3. Answers - here the practitioner’s will summarise the responses from each category on the explanation section (tab 1) and answer some further questions relating to the context;
4. Identified risks - this column will allow the practitioner to paraphrase and explain any risks identified from the explanation section relating to that question and/or category;
5. Privacy violation risk score - here the practitioner can rate the privacy risk high, medium or low, the fields will auto-colour; red, amber green depending on the choice made;
6. Impact risk score - this asks the practitioner to score the likely impact if the privacy of the data subject(s) is violated, again using the same scoring mechanism as for the privacy risk score above;
7. Mitigation strategy - this column provides the practitioner with an opportunity to note any mitigations strategies that, if implemented, may reduce the risk;

8. Post mitigation strategy if mitigation applied - here the practitioner can reduce the risk score accordingly should the identified mitigation strategy be implemented.

Tab 3 - Decision

This tab is where the decision is noted and the reasons for that decision. Further, practitioners will be asked to outline who will be responsible for publishing and updating the data and who will be accountable if the decision is challenged. The tab will consist of four columns as follows:

1. Step - this will be broken into numbers corresponding with the group of questions;
2. Contextual integrity - this will hold two groups of rows of questions; decision to publish and decision to not publish;
3. Decision - this will contain questions about the decision reached;
4. Answers - here the practitioner will be able to respond to the questions.

6.4 Step 2 - Peer Group Evaluation Outcome

Prior to applying CLIFOD in practice, the protocol asked that the questions devised be evaluated by a group of peers (see Section 6.2.5). The outcome of these evaluations were:

Solicitor

The solicitor that reviewed CLIFOD highlighted a few questions that could be clarified better, these were amended. Other than that, the solicitor did not feel there were any areas within the questionnaire that had been omitted or that could leave a public body open to challenge.

Practitioner

Upon review, the practitioner made no suggestions for change to the questions. Further, the practitioner believed that the questions were relevant, applicable and had potential to work in practice.

Academic

The reviewing academic made suggestions with regards to the wording, grouping and formatting of some of the questions as well as the presentation and layout. As a result, some of the questions were revised, moved and reworded in accordance with the suggestions made.

Further, in the original CLIFOD questionnaire, the first phase had been renamed to 'Fact'. However, based on the responses received from the expert reviewers, this was revised back to the description Nissenbaum uses for CI ('Explanation') as the consensus appeared to be that changing this was unnecessary. In addition, as a result of the expert comments and suggestions, some additional questions were added and the wording of some of the questions were amended for clarification. The questions in Appendix 6.2 represent the final questionnaire, including the changes made as a result of the expert reviews. For those interested, the first iteration of the CLIFOD questions sent to the reviewers can be found in Appendix B, Section B.3.

6.4.1 CLIFOD Layout Changes

Based on feedback from the experts some changes were also made to how CLIFOD is presented to practitioners. CLIFOD will still be presented as a spread sheet with three worksheets, one for each phase (see Section 6.3.8). However, on the risk assessment worksheet within the spreadsheet, the layout was changed as follows:

Worksheet 2 - Risk Assessment

On the risk assessment worksheet (a.k.a. 'tab') an additional row was added at the top of the worksheet to include an explanation of what information to note within each column as follows:

Applying CI framework to dataset - Where there is more than one attribute field within a group for a particular category, please insert additional lines to allow for each set;

Identified risks - Note all risks that may be applicable, where multiple risks identified, please insert a line for each risk;

Privacy violation risk score Please rate the risk of the privacy of the data subject(s) being violated;

Impact risk score Please rate the likely impact of the risk identifying on the data subject(s);

Mitigation strategy Please note any mitigation strategy that might be applied to minimise or eliminate the risk;

Post mitigation strategy if mitigation applied New score after/if mitigation strategy applied. If it is determined that no mitigation strategy is available or applicable this score will be the same as the original risk assessment score.

Worksheet 4 - Definitions-Key

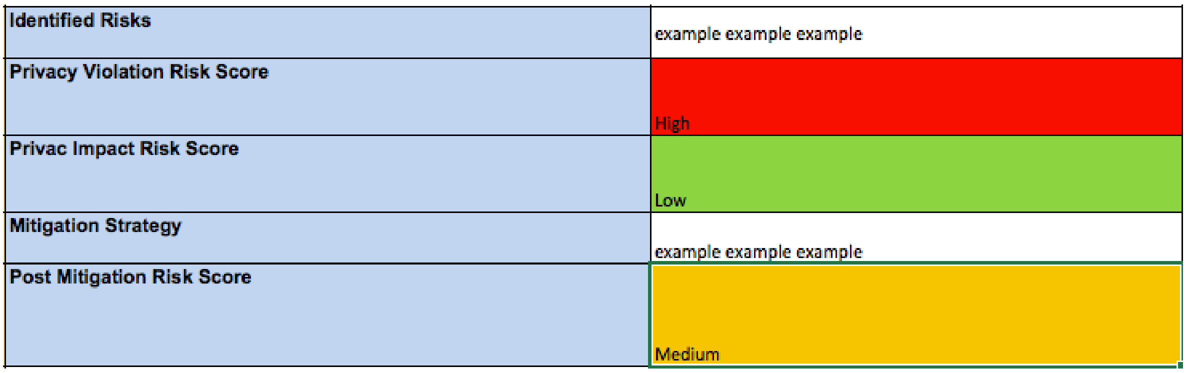
An additional worksheet ('tab') was added to include definitions for the terminology used. The definitions were:

Personal Data The Data Protection Act 1998 (DPA) refers to personal data as: data which relates to a living person who can be identified (DPA, s. 1(1)). To distinguish between data that aids re-identification or not, attributes can be broken down into: *identifiers/personal identifiers*, *Quasi-identifiers*; *Sensitive Attributes* and *Non-sensitive attributes* (see Chapter 2, Section 2.4.3); and

Data Processing Under DPA, processing refers to any handling of the data, be that collecting, manipulating, storing, processing or any other handling of the data (DPA, s. 1(1)).

6.4.2 Scoring

As part of the planning for CLIFOD, a decision was made to score practitioner responses by using a simple traffic light style marking scheme of low, medium, high similar to that used project and risk management practices (see Appendix 6.2, Section 6.3.7). This means that, as part of completing CLIFOD, practitioners will be asked to rank each answer or group of answers in accordance with the level of perceived privacy risk for that item as they work through CLIFOD. These scores will then be gathered together at the end of the risk assessment (phase two) so that the assessing practitioner will have an overview of where potential conflicts might arise. These scores or rankings can then be used to assess potential impact. Practitioners are then asked to consider and note any mitigation strategies and then re-score taking any mitigation applied (or to be applied) into consideration. An example of the risk score overview can be found in Figure 6.1.



Identified Risks	example example example
Privacy Violation Risk Score	🟥 High
Privac Impact Risk Score	🟩 Low
Mitigation Strategy	example example example
Post Mitigation Risk Score	🟨 Medium

Figure 6.1: CLIFOD - Risk Scoring

6.5 Step 3 - CLIFOD Practical trial

Once CLIFOD has been devised and evaluated by the peer group, the next phase will be to test whether CLIFOD can be effective as a privacy-decision support tool for making privacy decisions in regards to open data publishing.

On the day of the trial, the practitioners had prepared three datasets that had already been published for assessment using CLIFOD. The datasets chosen by the practitioner for assessment through CLIFOD were:

1. **Business rates** - Dataset derived from business rates paid by business in the LA area. This data forms part of a national dataset collected by each LA of rates paid by businesses within that LA;

2. **plosprimary** - School catchment choices. This is a dataset containing details of all applicants in past primary school admissions rounds - i.e. applications received for [LA] children starting primary, infant or reception;

3. **Section 106 Agreements** - Dataset containing a summary of all Section 106 Agreements, detailing financial contributions agreed between developers and the LA as part of the planning application process.

These datasets were all datasets that have already been published in open format. When asked why they did not select datasets in accordance with the agreed strategy (see Section 6.2.5), the reason given for this was that the Management team within the LA had determined that assessment of any datasets that have not as yet been deemed suitable for publication could pose a potential privacy risk. This was despite the fact that the researcher had providing assurances that first, the researcher would not need access to the dataset itself, only to any information discussed as part of the assessment itself. Second, the assessment would concentrate on the type of data captured within the dataset rather than individual records. Third, any information shared with the researcher would not be made available to anyone outside the research team and any data collected would be anonymised. Therefore, the assessments conducted were all carried out on existing datasets that had already been published.

Two practitioners and the researcher took part in the trial which took three hours to complete. The practitioners were given access to the CLIFOD questionnaire via Google Docs and, simultaneous to this, as part of the trial, the discussion around the assessment itself was conducted via telephone conferencing link. The results highlighted a number of privacy concerns for all of the datasets assessed, details of which are described in more detail in Section 6.6.

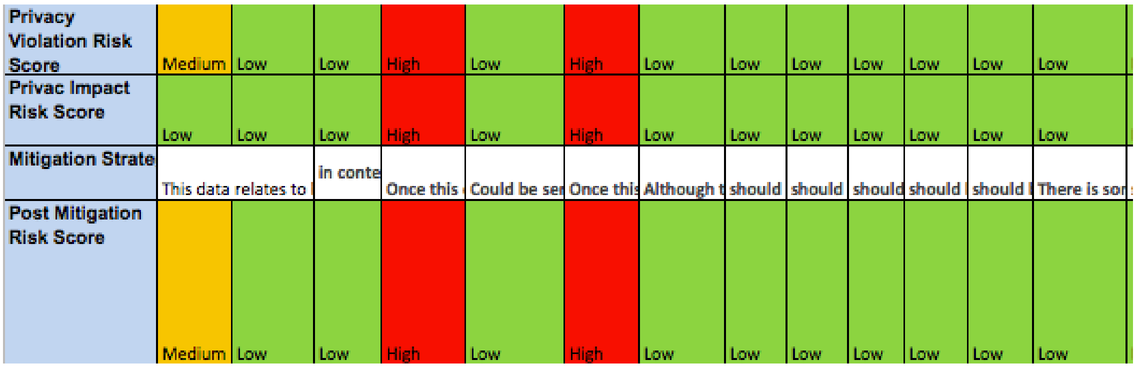
6.6 CLIFOD Findings

The CLIFOD assessment commenced with the first dataset, the business rates, and it quickly became apparent that not enough detail had been provided in the questionnaire to explain the process. This meant that a more detailed explanation of the concepts and reasoning behind some of the questions proved necessary to provide practitioners with clarification and encourage them to provide fuller, more thought out responses.

Once these discussions had taken place however, the participants appeared to start to appreciate the complexity of assessing privacy and really began to think about the data in context, making the rest of the assessment much more insightful for everyone. Furthermore, as the trial continued a number of other issues and logistical challenges presented themselves as part of going through answering the questions within CLIFOD for each dataset, these are outlined and explained further in Section 6.6.4.

6.6.1 Dataset 1

Based on the assessment, the outcome or privacy scores for this dataset showed that two of the attributes were classed as high risk and one attribute medium risk while the remaining attributes were all scored as low risk. These scores can be found Figure 6.2 below.



Privacy Violation Risk Score	Medium	Low	Low	High	Low	High	Low	Low	Low	Low	Low	Low	Low
Privac Impact Risk Score	Low	Low	Low	High	Low	High	Low	Low	Low	Low	Low	Low	Low
Mitigation Strate	This data relates to		in conte	Once this	Could be ser	Once this	Although t	should	should	should	should	should I	There is sor
Post Mitigation Risk Score	Medium	Low	Low	High	Low	High	Low	Low	Low	Low	Low	Low	Low

Figure 6.2: Dataset 1 - Privacy Risk Score

The main finding on assessing this dataset, the business rates, was that the data included identifiers (see Section 2.4.3), this was the reason for the score of "high". Upon discussing these scores, for one of the attributes marked high risk the practitioner explained: *"this attribute contains a national system reference number - this does contain a pattern, the first section will be allocated to a particular LA, the second part is a unique identifier for that particular reference. So this is a unique national identifier for that property that may give indication of where in the country it is and who/which business it relates to" (P1)*. The other practitioner went on to explain that: *"once this element of data is released, there can be no safeguards or guarantees that this element will not also be present with* *personal information. For example, if I was to associate this with my address that would not be a problem but if another dataset was to be released that contain this field and some personal information, that could then make this a very sensitive field" (P2)*.

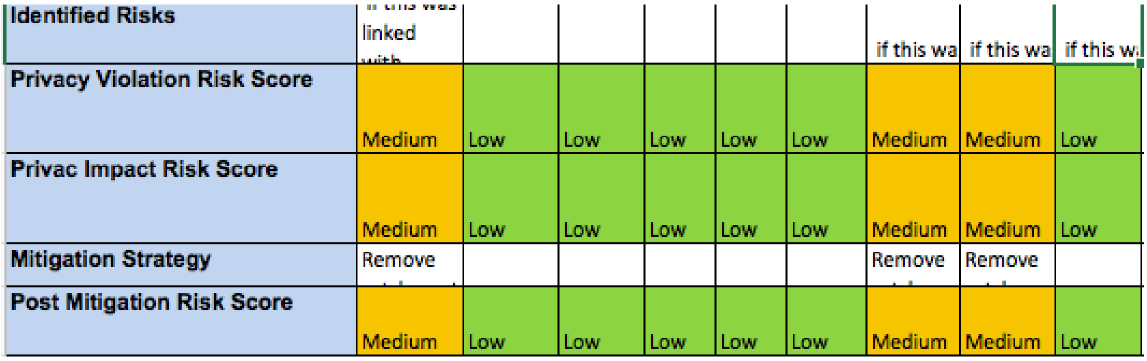
The reason these attributes were still included in the open data release, was because it related to businesses rather than private individuals and: *"some businesses may contain personal names if for example, the business owner uses their name as the business name as well" (P2)* and, in such instances, this information will be included in the dataset. However, the other practitioner further qualified this by explaining (in the mitigation strategy column on the risk assessment worksheet): *"this data relates to businesses rather than individuals and, as such, most businesses have a legal obligation to divulge this information so therefore, it cannot be considered personally identifiable in relation to individual people or personal data question" (P1).* For that reason, the dataset had been deemed suitable for publication in open format despite containing potentially identifying data.

When it came to answering the questions that related to the actors, the data controller and processors, the practitioners were not able to identify which employees had worked on the data beyond knowing that it would be employees within the relevant department. Therefore, for most of the questions relating to the actors and the context, the questions were answered rather vague with either "multiple" without further elaboration. For example, for Q30-42, which seek to gather information about the relationships between the data controller and/or data processors and the data subject, the practitioners were unable to provide detailed answers. As a result, a generic answer for the dataset was supplied stating: *"The relationship is that of debtor/client, it is professional only - even if there is some potential of the two being related this would not affect this. The rate values are pre-set by central government" (P1)*. Therefore, perhaps some further thought needs to be given to whether the practitioner responsible for publication is best placed to answer these questions.

6.6.2 Dataset 2

In the second dataset, the school catchment dataset, no directly identifying information was found to be present. However, a few of the attributes were scored as "Medium" risk (see Figure 6.3). In assessing this one of the attributes within the dataset confirms which school a child has been allocated to by a catchment code. This catchment code attribute, was classed as non-identifiable in Q1 (identifiers, see Appendix 6.2, Section 6.3.3) as the code does not provide sufficient detail in itself. However, in considering Q4 (risk of attribute being identifiable through inference), the attribute was noted as potentially sensitive.

The reason that this attribute was considered sensitive in Q4 was because it contained a pattern. The practitioner explained: *"this indicates which school a child has been allocated to by a catchment code, which is included in the dataset" (P2)*. Were an informed



Identified Risks	linked						if this wa	if this wa	if this w
Privacy Violation Risk Score	Medium	Low	Low	Low	Low	Low	Medium	Medium	Low
Privac Impact Risk Score	Medium	Low	Low	Low	Low	Low	Medium	Medium	Low
Mitigation Strategy	Remove						Remove	Remove	
Post Mitigation Risk Score	Medium	Low	Low	Low	Low	Low	Medium	Medium	Low

Figure 6.3: Dataset 2 - Privacy Risk Score

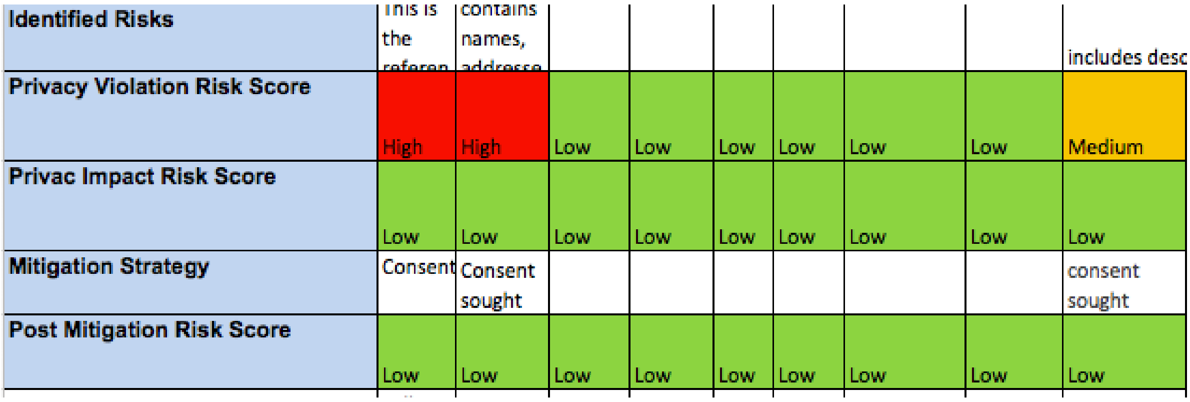
user to look at this, they would be able to identify which school the child has been allocated to. The practitioner continued: *"while it is not possible to identify the child from this, it is possible to work out which school the child is going to by name" (P2)*.

Similarly, for the same attribute, in Q74, (risk if data linked to external data), it was noted that there is a risk of re-identification, however, the practitioner considered this risk low stating: *"it could be possible with some prior knowledge and other data to work out which school a particular child has been allocated to but the risk is considered low given that this information can be gleaned in an easier way from other sources" (P2)*. Therefore, the practitioner has effectively confirmed that if this was linked with some knowledge of the family or child such as where they live, it would be possible to utilise this attribute as a quasi-identifier or sensitive attribute and potentially re-identify the family and therefore the child.

6.6.3 Dataset 3

The third dataset, the Section 106 dataset (S106), was also found to contain identifiers. This dataset contains a: *"summary of all current financial contributions arising from Section 106 Agreements in the local planning authority area. Section 106 contributions are made by developers through the planning application process and are mainly used to fund improvements to school, highway, green space and recreation infrastructure required because of the development" (S106)*. The score for this datasets can be found in Figure 6.4.

When asked why some of the attributes in this dataset were scored as "high", the practitioner explained that not all planning applications will originate from developers, some will originate from ordinary individuals who wish to build upon their land or make improvements to an existing property. For example, in the dataset under assessment, a number of entries referred to what appeared to be private individuals (giving names and addresses) who were seeking to make such improvements but who, as part of the agreement, had been ordered to make *"a contribution for off-site provision of open space"*



Identified Risks	This is the referen	contains names, addresse							includes desc
Privacy Violation Risk Score	High	High	Low	Low	Low	Low	Low	Low	Medium
Privac Impact Risk Score	Low	Low	Low	Low	Low	Low	Low	Low	Low
Mitigation Strategy	Consent	Consent sought							consent sought
Post Mitigation Risk Score	Low	Low	Low	Low	Low	Low	Low	Low	Low

Figure 6.4: Dataset 3 - Privacy Risk Score

(S106). However, for those same attributes, the privacy impact risk score had been set to ”low”.

When asked about this, the practitioner explained that planning applications are public record: *”accessible for public inspection either in person or online. These will include decisions made, plans, comments (anonymised) etc.” (P1)*. For that reason, according to the practitioners, when applicants submit an application, they will be aware that the information they provide may be shared with interested third parties. This, according to the practitioners, meant that consent had been sought and granted, and therefore, explained why, although some attributes were initially scored high, the privacy impact score had been set to ”low”.

6.6.4 CLIFOD Design modifications

Returning to the design of the questionnaire, in addition to the observations noted in the sections above, a number of changes will also need to be made in future iterations of CLIFOD to improve the flow of the questionnaire and the logic of how questions are grouped. The overarching finding of how CLIFOD flowed was that questions and answer sections would flow better if they were also grouped within each phase in relation to which key element they consider, i.e. attributes, actors or data flow. In the sections below, these are addressed in turn and been listed in accordance with their question number(s) (see Appendix 6.2, Section 6.3.3) in the sections below.

Pre-Assessment info - Describe Dataset

One of the first challenges encountered, occurred at the beginning of the trial when describing the attributes within the dataset being considered. This challenge concerned the number of attributes within the dataset. The first dataset being assessed was the Business Rates dataset (BR). This dataset contained twenty-two different attribute types, each of which needed assessing against all of the questions. However, the way CLIFOD had been designed and presented, while it asked how many attributes and asked for these to be listed, CLIFOD did not accommodate answering each question for this many attributes. To address this the format of the CLIFOD questionnaire needed altering to better accommodate the answering process. This instigated a discussion as to how best to present this as, going through all the questions 22 times did not seem like a logical or feasible option to the practitioners or the researcher. The solution devised was to add additional columns so that a separate column for each attribute group/type was created. Then in answering each question, answer that same question for each of the attributes across the columns at the same time to avoid having to go over the same question multiple times, i.e. effectively answer each question 22 times before moving to the next question.

Q1 - Personal Identifiers

This turned out to be a yes/no answer for all the attributes. Where a no was recorded, considering who was involved, how they were involved etc. did not appear relevant. Therefore, initially the practitioner commented: *"it might be helpful if, for this attribute, most of the rest of the questions within the explanation phase for that attribute were greyed out" (P2)*. However, a bit further on during the assessment, it was acknowledged that, at least for the risk assessment and decision phases, the attribute should again be included so that the questions in those phases could be answered for that attribute as well.

Q4 - Quasi-identifies

This discussion around the format of CLIFOD and greying out of sections was elaborated on during the second run through of CLIFOD when assessing the second dataset. What happened was that, as part of this second assessment, a yes was entered for Q4 (quasi-identifiers) on an attribute that had attracted a 'no' for Q1-3 (identifiers). This led to a conversation about the discussion on what aspects should be greyed out and when and culminated in the practitioner conceding this would need some further consideration and thinking about.

Q5 - Attribute description

In answering Q5, *"Describe the attribute, exactly what data is collected"* it turned out that, as part of answering that initial question, the practitioners had actually pre-empted many of the questions that followed and answered them as part of this first question. Therefore, in revising CLIFOD, these questions can be amalgamated or reworded to reflect this better.

Q1-Q12

For questions Q1-Q12, rather than grey out the rest of the questions in the explanation phase as suggested by the practitioner, this needs considering in a bit more depth. For example, perhaps where a 'no' answer has been entered for each of the questions within that grouping can be greyed out, with the groupings being as follows: Q1-3, Q4-6, Q7-9 and Q10-12, i.e. for Q1-3, if a no is entered, grey out questions 2-3.

Q14-15 - Attribute description

Q14 asks whether the data is original to the dataset being assessed or whether it was collected from another (i.e. is it owned or part of a third party or other dataset). Q15 then is intended to elaborate on this if the answer to Q14 is yes. The practitioner correctly pointed out that this surely should be if the answer is no. Therefore, this question should be rephrased to read *"If NO, who/which department does the data originate from"*.

Q17-Q47

The answers to questions 17 to 47 all concern the actors (or 'users') both internal; i.e. the data- owner, controller and processors and external; i.e. the data subject and the end user or data recipient. Therefore, these questions will only need answering once for each dataset. However, because there may be multiple internal users, and the questions that follow will therefore need to be considered for each internal user. To allow for this as part of the trial, it was agreed that a column should be created for each actor so that the answers for that actor could be recorded within each column. However, some thought will need to be given to how future iterations of CLIFOD can make this transition easier. It may be that the actor and attribute gathering needs separating out and grouping in a more logical format within the explanation phase before amalgamating these back into the mix for Phases 2 and 3.

Q30-Q42

As mentioned in Section 6.6.1, the practitioners did not have sufficient knowledge of who had worked with the data prior to publication and therefore future work will need to give consideration to potential process changes for applying CLIFOD in practice. For example, if CLIFOD was to be split so that different actors within the LA could consider different aspects within their area or expertise, would that make the privacy risk assessment more accurate? Alternatively, perhaps multiple internal users (actors) should be asked to complete the CLIFOD assessment in isolation and then a designated privacy aware person should be asked to make the decision based on a group of assessments of the same dataset.

Q48-Q54 - Attributes

These questions again pertain to the dataset rather than the actors and therefore, these should be grouped with Q1-12 above rather than with the actors.

Q55-Q68

These questions pertain to the actors and what permissions were granted as part of the original data collection rather than the dataset itself. Therefore, these need to be grouped and considered with the actors.

Risk Assessment Q88

At the end of the risk assessment questions (after question 88), the practitioners suggested that another question should be added to allow them to account for any 'other' risks not specifically mentioned or identified as part of the assessment. For example, for the BR dataset, they wanted to include a risk concerning a potential financial risk as, the practitioner explained: *"businesses could use this information to gain financial advantage. This risk is seen as medium to low. Medium referring to a potential financial risk as in revenue losses for the individual" (P1)*.

Other Modifications

Looking at the general presentation of CLIFOD and how this is conveyed to practitioners, in future versions of CLIFOD it might also be useful to:

a Provide more detailed explanation for each section to include more of the detailed outline in devising the questions originally (see Appendix 6.2);

b Include more definitions section within the questionnaire, either on the definitions worksheet and/or within the answer worksheets themselves;

c Split the assessment to allow multiple practitioners to complete different aspects of the questionnaire. For example, it might be useful if the data owner were to be involved in the CLIFOD assessment. Although the practitioners responsible for publication assess privacy, the ultimately responsibility for the data lies with the data owner, and therefore, it may be beneficial if they were to conduct all or part of the CLIFOD assessment (see Section 6.3).

6.7 Summary of Findings

In this Chapter, the meta-model developed in Chapter 5, was used as the starting point for creating a paper prototype: the CLIFOD questionnaire that will form the basis for a theoretically grounded decision-making framework.

The meta-model demonstrated that an existing framework can be adapted to provide a practical foundation for privacy risk identification and, in doing so, served as a proof of concept for creating a privacy-specific theoretically grounded decision making framework. Based on this, CLIFOD was created to show how this can be adopted to support privacy decision-making for making data available in open format.

CLIFOD was then applied in a case study where the privacy implications of previously released open data was assessed using CLIFOD. This trial was conducted in a real setting, using real data working in collaboration with a public body, assessing real data that has already been released as open data.

Three datasets were assessed using CLIFOD and the findings were that all of the datasets contained varying degrees of potentially sensitive information. The sensitivity of the data released in open format covered the full spectrum with one dataset containing direct identifiers (see Section 6.6.3), while the other contained quasi-identifiers (see Section 6.6.2), or sensitive attributes (Section 6.6.1). These findings it can be clearly demonstrated that using a privacy decision framework such as CLIFOD, could help prevent sensitive data being released in open format and thus, preserve user confidentiality and privacy.

The CLIFOD framework is an exemplar tool for effectively supporting privacy decision making. It demonstrates how Contextual Integrity, can be adapted to create a privacy specific decision framework that can be applied in practice by practitioners to support privacy decision making in the domain of *public open data*.

In doing so, CLIFOD answers RQ1 (see Chapter 1, Section 1.2) by providing a useful prototype tool that facilitates privacy decision making which can be applied in practice. Further, in light of the fact that the *public open data* domain is particularly problematic as data is made freely available with no control or restrictions, it is contended that these concepts and thus, the CLIFOD framework will work equally well in supporting privacy decision making for other domains as well.

6.7.1 Next Steps

The next step will look at how contemporary legislation can be incorporated into the CLIFOD framework and establish whether the contention that the concepts and questions created for the *public open data* domain can be equally applicable in other industry sectors (domains) where data is processed and shared. This will be done as part of a case study and trialled working in collaboration with another domain, the charity sector.

Chapter 7

DPIA Case Study

7.1 Introduction

In Chapter 6, Contextual Integrity (CI) was used to create a privacy decision making framework for determining the suitability of data for open data publication. This framework was successfully trialled in practice working in collaboration with a local authority (see Chapter 6, Section 6.6), thereby demonstrating how using a risk based approach combined with CI can be used to facilitate effective privacy decision making in the *public open data* domain. This chapter will expand on this work by incorporating the idea of CI into other areas where privacy decision-making is needed within organisations and trialling this in other domains. In particular, this chapter will support and facilitate conducting Data Protection Impact Assessments (DPIA), a particular type of privacy risk assessment that seeks to assess privacy risks to the individual when processing personal data.

The reason for extending the work into this area is to answer RQ2 which asks: *"How can contemporary legislation be incorporated into the privacy-specific assessment framework (PAF) to practically support practitioners in privacy decision-making?"* (see Chapter 1, Section 1.2). To this end, the General Data Protection Regulation (GDPR) brought in by the European Union (EU) in May 2018 (European Parliament and the Council of Europe 2016) and implemented into UK law under the Data Protection Act 2018 (UKDPA). UKDPA was enacted on 23rd May 2018 to implement the provisions of GDPR into UK law (UK Parliament 2018). These provision closely follow the GDPR but there are some exemptions, restrictions and adaptations in the UKDPA in sections 15 and 16 (UK Parliament 2018). However, at the time of conducting this study, the final wording and provisions of UKDPA had not as yet been published and therefore, the advice and guidance produced as part of this study has been conducted and guided only by the provisions of the GDPR.

GDPR imposes a number of new obligations on organisations; these include extending the scope and breadth of what data is classed as personal, more rights for individuals in relation to their data; a requirement for organisations to understand and document their data holdings; justify why they collect each piece of data and record the lawful basis for processing data. In addition, organisations must also implement privacy by design and default, and be able to demonstrate compliance to the relevant authorities if challenged or a breach occurs (see Chapter 2, Section 2.4.4). Thus, this new legislation provides an excellent opportunity to answer the research question.

This work was designed to follow on from the meta-model and CLIFOD (see Chapters 5 and 6) and build on these models to create an accurate view of what data holdings the Charity has and develop the DPIA framework. This will be done as an action intervention case study in alignment with the main methodology (see Chapter 3, Section 3.5).

This rest of this chapter is organised as follows. The Chapter starts in Section 7.2 by providing the case study protocol for this action intervention. Section 7.3 describes what happened in the case study, explaining each of the phases starting with Phase one, where the Charity's data holdings are assessed (Section 7.3.1). Next in Phase two, the data holdings and processes are analysed (Section 7.3.2) which involves creating three prototype spreadsheets: the Staff User Stories; the Life of the Form; and a Master Data Register (MDR) for the Charity. This is followed in Phase 3, by a section on GDPR processes, including guidance on what step must be taken to bring these in line with GDPR (Section 7.3.3). Next in Phase 4 (Section 7.3.4), a Data Protection Impact Assessment (DPIA) is created and evaluated to facilitate the Charity conducting standardised DPIAs. The Chapter concludes with a summary of findings in Section 7.4.

7.2 DPIA Case Study Protocol

This case study seeks to address this problem by providing an exemplar approach to GDPR implementation that incorporates the design and creation of a DPIA framework, aimed specifically at the charity sector. Working in collaboration with a local charity, this study presents a action intervention for how GDPR was implemented within the charity, providing them with a data register and a framework for conducting repeatable, consistent and compliant DPIAs going forward.

The charity works with addicted persons helping them overcome problems of substance and/or alcohol abuse. As such, they collect and process a lot of very personal information as part of their work. As part of the treatment the charity shares client data with external stakeholders such as clinicians and professionals who may provide treatment as part of the charity's or the clients support package. In addition, other stakeholders that the charity may share data with include national addiction monitoring bodies, such as the National Drug Evidence Centre (NDEC), part of Public Health England (National Drug Evidence Centre 2018), who collate statistical information on adult addiction users and other Governing bodies such as the Care Quality Commission (CQC) who regulate health and social care in England (Care Quality Commission (CQC) 2018).

The other problem that the Charity face is how to deal with consent. Under the data protection legislation, both the Data Protection Act 1998 and the new General Data Protection Regulation, the Charity must seek the client's (data subject's) consent for the processing of personal data (see Chapter 2, Section 2.4.4). However, as the Charity deals with vulnerable clients, who suffer problems with addiction and abuse, there is a risk that the clients' circumstances can change while they are being supported or undergoing treatment, resulting in them either not being able to provide consent or claiming that they were not capable of providing consent at the time of entering treatment as they lacked mental capacity. This could undermine the Charity's ability to demonstrate compliance under GDPR as the Charity presently relies on client consent in order to provide effective treatment to their clients. However, GDPR requires that clients must be able to withdraw their consent at any time. This imposes a risk that any services provided that is reliant on consent can no longer be provided effectively or, if treatment continues, could result in the Charity potentially contravening GDPR. There is no easy solution to this problem. However, by careful planning, it is possible for the Charity to safeguard themselves and their clients to address this, this is discussed in Section 7.3.3.

While GDPR may not necessarily require expert knowledge to implement, the requirements and obligations still require interpretation and charities, like many organisations, were finding it difficult to fully understand how best to implement GDPR and conduct Data Protection Impact Assessments (DPIAs). The Information Commissioners Office (ICO) has provided some information and general guidance on GDPR and how to implement the regulation, made available through their website (ICO 2017 2018b). However, this guidance is generic and not directed to any particular industry sector. Charities with their unique qualities and requirements have requested that more sector specific guidance be created for the Charitable sector but, in view of the large workload of the ICO, so far, this has not been forthcoming (ICO 2018c).

7.2.1 Problem Definition

The General Data Protection Regulation (GDPR) is the European Union's (EU) new Data Protection Regulations that take effect on 25th May 2018. In the UK, GDPR will take effect from May 2018 despite the fact that the UK is in the process of leaving the EU (Financial Conduct Authority 2018). Thus, GDPR will affect all organisations, but has particular implications for charities, who, like many other organisations, collect and process personal data. Having a detailed GDPR implementation plan and conducting Data Privacy Impact Assessments (DPIAs) helps organisations evaluate their readiness and ability to demonstrate GDPR compliance to the ICO. While GDPR may not necessarily require expert knowledge to implement, the requirements and obligations still require interpretation and many organisations struggle in fully understanding how best to conduct DPIAs and demonstrate compliance.

Most charities rely on public generosity for funding and in-kind support from volunteers to function and, as a result, many charities struggle to raise enough funding to meet all the objectives for their cause. Further, charities, as part of having charitable status, must be accountable and transparent in how they spend any money raised (Gov.UK 2016b). This means that charities often face conflicting demands on how best to spend funds received and deciding which issues take priority. Further, because of the nature of charitable work, a lot of the work is conducted by volunteers meaning that, even though a charity may collect and manage personal data, they often lack the resources and expertise to assess themselves against regulations such as GDPR.

This action intervention case study seeks to help address the problem of implementing GDPR and conducting data privacy impact assessments (DPIA), working with a charity that provides advice, support and information to vulnerable people in the local area. The charity supports those suffering from addiction and substance misuse. As such, this charity needs to collect personal information from clients in order to provide them with the care, assistance and help they need for dealing with or overcoming their problems.

There are no current solutions for implementing GDPR or carrying out DPIAs in this context and therefore, this work will benefit not only this charity, but other charities with a vulnerable client base. Thus, this work will be conducted as a case study (see Chapter 3, Section 3.4.5), designed to follow on from previous work that took Nissenbaum's Contextual Integrity (CI) Framework (Nissenbaum 2010) to create:

Meta-model : a meta-model for how to assess privacy risks in Open Data;

CLIFOD : Contextual Integrity for Open Data in Practice (CLIFOD), a privacy risk assessment questionnaire for assessing privacy risks in releasing data for publication.

The intention is that this case study will build on these models to implement GDPR within the Charity. To help organisations implement GDPR, ICO have published a twelve step guide in how to prepare for GDPR (ICO 2017) which states organisations should:

1. **Awareness**: ensure that key stakeholders are aware of the change in the law brought by GDPR;

2. **Information**: seek to document their data holding by conducting an information audit;

3. **Communication**: Review and update privacy notices;

4. **Individual rights**: ensure they have procedures in place that adheres to individuals' rights, including any request for providing or deleting personal data;

5. **Subject Access Request**: ensure, as part of access request processes, they are able to meet and respond to such requests within the permitted time scales;

6. **Lawful basis for processing personal data**: identify and document under what lawful basis they are collecting or processing data and update privacy policies accordingly;

7. **Consent**: review current processes for obtaining consent and make any necessary changes to comply with GDPR;

8. **Children**: assess whether any personal data collected or processed relates to children and, if so, ensure appropriate safeguards are put in place to verify client's ages and/or seek parental or guardian consent for any processing;

9. **Data Breaches**: create appropriate processes and procedures for detecting, reporting and investigating any data breaches involving persona data;

10. **Privacy by Design (PbD) and DPIAs**: familiarise themselves with how to implement data protection by design (a.k.a. Privacy by Design) and conducting DPIAs;

11. **Data Protection Officers**: designate an organisational Data Protection Officer, tasked with ensuring organisational compliance with GDPR;

12. **International**: where the organisation carry out any cross-border data transactions or processing, establish who is to act as lead data protection authority.

In the above list, data refers to any data or information the organisation collects, stores and/or processes that may, or could, be considered either personal or sensitive, i.e. personally identifiable information such as names, addresses, date of birth etc.

Thus, for our charity, they will need to carry out a number of actions in order to effectively implement and embed GDPR within the organisation's processes. However, like many charities, this organisation struggles with fully understanding how best to implement and comply with GDPR. They also find it difficult to find sufficient resource and time to demonstrate compliance with the regulation by May 2018.

The other problem that the charity faces concerns consent. As required by the data protection legislation (both the current Data Protection Act 1998 and the new General Data Protection Regulations) the charity seeks consent from clients for collecting, processing and storing their personal data. However, clients' circumstances may change during the course of treatment or receiving support and assistance that would necessitate obtaining consent again for additional processing or handling of data at which points clients' may not be able to provide such consent due to the effects of addiction. Alternatively, because of the effects of addiction, the charities' clients may later claim that they lacked the mental capacity to provide informed consent when they went in for treatment which will undermine any demonstration of GDPR compliance, particularly when it comes to ensuring data subjects can exercise their right to be forgotten. The project will, time permitting, also look at how this issue may be addressed for the Charity.

The aim of this action intervention case study is to assist a UK Charitable Organisation in implementing the General Data Protection Regulation (GDPR) by guiding them through the GDPR implementation process and provide guidance on how to conduct DPIAs going forward. Further, the study also aims to provide the charity with pointers to suitable tool-support that they can utilise in managing both the process and meta-data collected during the assessment. This work will be done as a case study, in line with the main, overarching methodology (see Chapter 3, Section 3.5).

For this, a number of steps will need to be carried out starting with gaining an overview of what data the Charity hold and in what format so that an assessment can be made of the Charity's GDPR readiness. To this end, user stories will be collected to gain understanding of how staff and volunteers currently use and handle data as part of their daily work. The data gathered as part of this will then be used to create a data register of the Charity's data holdings and inform the remainder of the study.

7.2.2 Research Questions

The question this case study seeks to answer is RQ2 (see Section 1.2). However, in determining the protocol, it became evident that having some sub-questions and propositions relating specifically to the problem area, i.e. implementing GDPR and creating a DPIA process, would be beneficial for clarity. Therefore, this action intervention case study will seek to answer the following sub-questions:

CS-RQ1 What data holdings does the Charity have, where and how are these handled currently and to what extent do these comply with GDPR standards?

CS-RQ2 What processes does the organisation need to put in place for effective GDPR implementation to demonstrate GDPR compliance?

CS-RQ3 How can the organisation ensure they have in place appropriate processes conducting DPIAs going forward?

Propositions

The supporting propositions (P) for answering the sub-questions are:

CS-P1 Existing processes are not compliant with GDPR standards;

CS-P2 Current processes are inappropriate for supporting effective GDPR implementation;

CS-P3 The organisation does not conduct DPIAs as standard.

7.2.3 Ethics

Ethics approval for this case study was sought and granted from the University Ethics Committee, a copy of the approval can be found in Appendix C, Section C.1.

7.2.4 Methodology

The unit of analysis being ’the Charity practitioner’. The reason charities have been chosen as the collaborators is that charities are more vulnerable as they normally have to rely on the public for funding and are often run by volunteers who may or may not be conversant with legal regulations and how best to implement them. Therefore, charities may find it harder to raise enough funding to meet all the objectives for meeting new obligations imposed on them such as GDPR. Moreover, charities must be transparent and accountable in how funds are spent (Gov.UK 2016b) which can result in difficult decisions needing to be made over how funds are best distributed between the charitable work and supporting administration and which projects and initiatives take priority. Therefore, although a charity will likely collect and manage personal data, they may lack the resources and expertise to assess themselves against regulations such as GDPR.

In order to answer the research questions the case study will be carried out in four phases:

Phase 1 - Data Holdings Assessment : assessing what data the Charity’s currently collects, stores and processes and how this data is collected, handled, stored and shared (processed) this will take the form of collecting user stories (see Section 3.4.9). However, assessing the data holdings will require a series of questions to be asked as part if the user story collection. To ensure these questions are relevant, valid and correct a Pilot study will be conducted, the details of this Pilot can be found in Appendix C, Section C.2;

Phase 2 - Analyse: data holdings to create a Data Register : based on the findings from phase 1, a register of the organisation’s data holding will be created to gain an overview of what personal data the charity holds, where it originated from and who they share the data with;

Phase 3 - GDPR Process Guidance : an assessment will be made of how the organisation can best demonstrate compliance with GDPR principles in order to develop appropriate guidance as to what processes and actions the Charity will need to put in place to demonstrate GDPR compliance;

Phase 4 - DPIA : develop Data Privacy Impact Assessment (DPIA) process for data handling going forward.

Further details of each phase can be found in the sections below.

7.2.5 Phase 1 - Data Holdings Assessment

Phase 1 involves to creating an accurate view of what data holdings the Charity currently has. This will involve assessing what data holdings the Charity current have and how data is processed and shared both within the organisation and externally.

Relating this to previous work, in terms of the meta-model and CLIFOD, this equates to the explanation phase where we need to gather details about the data, the actors, their roles, and how the data is transmitted. However, in this instance, this does not involve assessing a particular dataset, rather, it involves piecing together an overview of what data the Charity holds and how the data is used, stored and shared. Therefore, this phase will consist of gathering information about what personal data the Charity currently collects, the format the data is collected in, how the data is processed and what stakeholders might utilise or access the data both internally and externally and how this process occurs.

To this end, phase one will begin by collecting user stories from staff and volunteers about how they currently use and process data and who they might share it with, using Storytelling as the research method (see Chapter 3, Section 3.4.9). This method has been chosen because of its simplicity. The intention is to get users to tell the researcher about their work and how, as part of this they handle data or information. This will allow users to recount their working day as a story in their own words which will provide valuable insight for the researcher into how data flows between stakeholders and what is captured where, when and how.

User Stories - data collection

To establish data flows and holdings within the Charity, users (or participants), in this case staff or volunteers who work with the charity, will be asked to provide their user story by completing a spread sheet detailing occurrences during a typical day where they collect, process, share or otherwise handle personal data as part of their daily tasks.

The spread sheet that participants will be supplied with will be in the same format and contain the same questions as described in the pilot study above.

The format of the questionnaire has deliberately been kept brief as the intention is to allow users to use their own words and phrases as they would when telling a story. The reason for this is firstly that this will provide the researcher with understanding how staff carry out the task being described. Second, this method will also providing valuable insight into what documents, forms, devices etc. the participants use as part of performing a task or completing a scenario that may otherwise be missed if too much has been pre-empted in the questions. However, to help users, two slightly different examples have been provided for this study. These have been typed in red italic font as to indicate they are merely examples and inserted in the first two rows of the spreadsheet.

The examples read as follows:

Example 1

1. **Tag**: 10 am meeting with client;
2. **Communication with/data accessed via**: Client;
3. **Physical location of user**: Office;
4. **Data handling method/Device used**: Manual, paper record & email (computer);
5. **Please make an entry for every time data is handled/processed/discussed during a typical day**: Met with AB, this was our first meeting. We went through the application form together and filled it in. Then, once the form was completed, both AB and I signed the form. I then gave AB a copy of the house rules. After the meeting, I copied the application form and filed the original in the filing cabinet in the common room. I then scanned and emailed a copy to admin and asked them to create an account for AB;
6. **How long does this scenario take?**: The meeting lasted 1.5 hours, I then spent another 20 minutes doing the paperwork;
7. **How often does this scenario occur?**: about 2-3 times each Month
8. **Please indicate how demanding you find this task, please explain your reasons**: This is quite a time consuming task and it can be emotionally draining as I have to be mindful the whole time of the client's feelings and needs. The admin aspect can be a bit of a cumbersome bind;
9. **Does this task interfere with any other tasks?**: When I am in a meeting with a client no-one usually disturbs us. The phone is put on silent and there is no computer in the meeting room. However, as soon as the meeting finishes and I get to the admin part, there tends to be constant interruptions.

Example 2

1. **Tag**: evening, 8 o'clock - checked emails
2. **Communication with/data accessed via**: Server;
3. **Physical location of user**: Home;
4. **Data handling method/Device used**: Home computer;
5. **Please make an entry for every time data is handled/processed/discussed during a typical day**: Logged into emails, there were two emails that I needed to

respond to. One concerned AB who had an argument with one of the other housemates. The case worker on duty had completed a complaint form. I downloaded the form and checked through it. I then signed it and uploaded the signed copy to the system;

6. **How long does this scenario take?**: About an hour;
7. **How often does this scenario occur?**: Every day
8. **Please indicate how demanding you find this task, please explain your reasons**: The task was not difficult, more disappointing because there had been a problem which will result in a lot of extra work for everyone involved. I will need to spend some time now doing reports etc.when I get into the office in the morning;
9. **Does this task interfere with any other tasks?**: Not really, I do not have to check emails from home but chose to so I don't arrive in the office the next morning to too many surprises.

User Story Data Collection

For the data collection the participants will be the any member of staff and/or volunteer who works and handles data as part of their work at the Charity. Therefore, to be effective, all staff and volunteers who handle data will need to take part in the study. To achieve this will require collaboration and assistance from the Management team within the Charity in order to collect the data. To this end the following format will be adopted for collecting the user stories from participants:

1. The researcher will email the Manager of the Charity and ask them to circulate the user story spread sheet together with a participant agreement form for them to complete. Attached to the email will be:
   (a) A participant information sheet for staff and volunteers (see Appendix C, Section C.5);
   (b) A participant agreement form (see Appendix A, Section A.3);
   (c) The user story spreadsheet.
2. The Charity Manager will email or circulate paper copies of the participant form and spread sheet to all staff members who process data asking them to complete the forms;
3. The Manager will collate the responses from staff and forward these to researcher within an agreed time frame to be agreed between the Charity Manager and the researcher.

7.2.6 Phase 2 - Analyse: data holdings to create a Data Register

This Phase involves analysing the data gathered from the user stories collected from staff and volunteers to produce an outline data register. This will involve analysing the data collected from users to created an outline data register of the Charity's data holdings. In addition, a list of documents and forms used by the Charity for gathering information will be collated and compared with information gathered from the user stories and added to the data register.

Forms and Documents - data collection

As part of phase 2, additional data will be gathered about what forms and documents the charity use for different scenarios. This will involve liaising with the Charity management team to collect copies of, or details of, all the forms and documents the Charity uses to collect and store information.

From these information about exactly what type of data is collected can be gleaned and added to the data register. This will allow the information pertaining to the Charity's data holdings to be mapped to the meta-model and analysed to create a Data Register for the Charity.

Data Register - data collection

The data register will be created based initially on findings from the user stories analysis. Once this analysis has been carried out details of the forms and documents used within the Charity will be added to the register. This will involve analysing exactly what data is collected, processed or shared within each document or form and categorising this data based on attribute types see Section 2.4.3 and data sensitivity. It may transpire as part of this analysis that not enough information is available and clarity needs to be sought. If this is the case, clarity will be sought through email exchange and/or semi-structured or contextual interviews see Section 3.5.2.

Data Register - Spread Sheet

The data register will be captured in a spread sheet initially with a view to load this information into a support tool, such as CAIRIS, at a later stage if time allows. As a starting point, the spread sheet will consist of the following columns and headings:

ID Each document or form will be given a unique identifier (*numerical*);

Date Document date;

FormName The name of the document or form (*text*);

DocType The type of document (e.g. template, form type such as application form, assessment etc.) (*drop-down list or text*);

DocFormat The format of the document (e.g. spread sheet, csv, office etc.) (*drop-down list or text*);

NoAtt The number of attributes (columns or data groups) captured on the form (*numerical*);

AttNo1....* A column will be created and assigned to each attribute based on the number of attributes captured in the previous column (with no maximum limit) (*text*);

AttType Each attribute will be categorised according to attribute type (i.e. identifier, quasi-identifiers, sensitive, non-sensitive) (see Section 2.4.3) (*drop-down list*);

Origin Where the form or document originates from (*text*);

NoUses The number of different uses the form or document has (*numerical*);

UseNo1....* A column will be created and assigned to each use based on the number of uses noted in the previous column (with no maximum limit) (*text*);

StorageMet Method of storing the form (if multiple, create multiple columns, one for each method) (e.g. electronic, paper)(*text*);

StorageLoc Location where the data is stored (if multiple, create multiple columns, one for each location) (*text*);

MinRetention Information about any minimum retention period applicable for storing the data (*text*);

MaxRetention Information about any maximum retention period applicable for storing the data (*text*);

Disclosure Details of who the data is/may be disclosed to (*text*);

This list is not exhaustive, more columns will be added as part of the analysis.

Concurrent with this analysis, a second round of data collection will be carried out to collect details of all forms and template documents created by the Charity in the course of their work. To create this list, the Management team at the Charity will be asked to provide the researcher with blank copies of all the template forms and documents used by the charity (i.e. forms that have not been completed). From this, a draft list of forms and documents will be created and added to the data register.

Data Register Validation

To validate the completeness of the list of forms, the following validation exercises will be carried out:

- The list will be compared with information provided in the user stories to ensure no documents or forms have been mentioned that are not on the list;
- The list will be emailed to the Management team, asking them to verify that no other templates are used or, if any are missing, provide copies to the researcher.

7.2.7 Life of the Form - data collection

Once the list of draft list of forms and documents has been verified, a third round of data collection will begin. This time to capture details of 'the life of the form'.

This will mean conducting another study to obtain details from users about how the use the forms and documents that have by now been collated. The format of this study's data collection will be the same as the format used for collecting the User Stories. However, the questions asked on the spread sheet circulated will be different.

Life of the Form - Spread Sheet

For this element, the data collection will start by taking an extract of the draft data register and creating a spread sheet with all the forms listed on it.

Users will then be asked to complete 'the story of the life of the form'. This exercise will be similar to the capturing of user stories only this time, users will be asked to provide details of:

Birth Where was the form born, who created it and why?

Intended uses What is the form intended for?

Actual uses What is the form actually used for?

Regularity How often is the form used?

Who Who fills in the form?

When In what circumstances is the form filled in?

Why Why is the form filled in?

Format What format is the form in (e.g. manual paper based or electronic)

Disposal How does the form get disposed of?

Journey What happens to the form next (this part is cyclical and applies every time the form travels) when:

1. It gets completed or it travels?
2. Where does it go?
3. Who is present?
4. Who sees the form?
5. Who can access the form?
6. How long does the form stay there?
7. Where does the form go next (this may be multiple places)?

Home Where is the form kept or stored (where does the form live)?

1. does it get moved from it's home (if yes, then back to the journey)?

StorageFormat In what format is it kept or stored?

Access Who can access the form once it is stored?

Retention How long is the form stored for?

Disposal How does the form get disposed of?

Like the draft data register list, this list of questions is not exhaustive.

Life of the Form - Validation

To validate the life of the form questionnaire, the following actions will be carried out before the questionnaire is sent to the Charity users for completion:

1. *GDPR* An analysis of GDPR data principles and standards will be carried out to ensure questions have been created that cover all aspects of GDPR principles within the questions asked;

2. *Pilot* The questionnaire will be piloted by asking 3 independent parties to complete it and comment on the clarity of the questions asked.

Once users have completed the 'life of the form' list of questions, all data gather will again be analysed and the data register will be updated. It may prove necessary, as part of this analysis, to seek further clarification. This may take the format of email exchanges between the researcher and users or, it may take the form of semi-structured interviews.

The outcome of this analysis will be the completed Data Register which will be handed over to the Charity for them to maintain going forward.

Data collected in Phase 1 and 2 will provide the basis for answering CS-RQ1: *What data holdings does the Charity have, where and how are these handled currently and to what extent do these comply with GDPR standards?* and, it is expected, confirm hypothesis CS-P1, that existing processes within the Charity are not compliant with GDPR standards.

7.2.8 Phase 3 - GDPR Process Guidance

Based on the findings from Phase 2, a guidance document will be created for the Charity in Phase 3. This document will outline what steps and procedures the Charity should consider putting in place to implement GDPR that will enable them to demonstrate compliance with GDPR.

To prepare this document, the following steps will be carried out:

1. Existing privacy policies to be reviewed and revised as necessary;
2. Review the Charity's existing processes for obtaining client's consent and revise as necessary;
3. Establish what processes exist for meeting data subject rights (GDPR, Chapter III);
4. Arrange staff training on GDPR.

This work will centre on identifying and helping the Charity prepare appropriate documentation and processes that will enable them to demonstrate GDPR compliance. The resulting report will outline what work has been carried out and identify any additional processes and procedures the Charity should consider putting in place to complete their GDPR implementation. Further, the report will also make recommendations as to what next steps the Charity should follow to be able to demonstrate compliance with GDPR. This report will provide the answer to CS-RQ2: *What processes does the organisation need to put in place for effective GDPR implementation to demonstrate GDPR compliance?* and, confirm or otherwise hypothesis CS-P2 that current processes are inappropriate for supporting GDPR implementation.

Further, in parallel with this work, the data register will be uploaded into CAIRIS so that work can begin on privacy threat modelling. Modelling privacy threats in CAIRIS, a security requirements gathering and threat modelling open source tool based on the IRIS security risk assessment framework Faily and Fléchais (2010b). By leveraging of this existing model which has been thoroughly tested and validated, the research team will be able to build on the security risk and threat modelling framework already incorporated into the CAIRIS system. This will help identify security risk as well as privacy risks and thus, aid in identifying areas where the Charity can make security improvements that will enhance their ability to demonstrate GDPR compliance.

Privacy Policy Review

As part of the process guidance review, the Charity's privacy policies will be reviewed and revised to ensure they align with the GDPR regulations. GDPR requires privacy policies to be "concise, easily accessible and easy to understand" and to this end, any policy must be written in "plain language" (GDPR, Recital 58).

7.2.9 Staff Training

Another aspect of GDPR implementation is to ensure staff are aware of their responsibilities as regards handling client's data and to this end some staff training sessions will be arranged. This will serve two purposes, first, it will enable staff to be informed of the changes brought in by GDPR and second, it will serve as an opportunity to evaluate the DPIA framework.

7.2.10 Phase 4 - DPIA

In Phase 4, a DPIA will be developed based on the outcome of Phases one to three. Further, the meta-model and CLIFOD will be utilised to inform the design of the final DPIA framework (see Chapters 5 and 6).

This DPIA will be trialled in collaboration with the Charity following the Case Study protocol utilised in the CLIFOD Case Study (see Appendix 6.2).

It is expected this phase will confirm CS-P3, that the Charity does not currently conduct DPIAs as standard.

The DPIA produced for this case study will be uploaded into CAIRIS and shared with the charity for their use going forward. This phase will provide the answer to CS-RQ3: *How can the organisation ensure they have in place appropriate processes conducting DPIAs going forward?*.

It is expected this phase will confirm CS-P3, that the Charity does not currently conduct DPIAs as standard.

The DPIA produced for this case study will be shared with the Charity for their use going forward. This phase will provide the answer to CS-RQ3: *How can the organisation ensure they have in place appropriate processes conducting DPIAs going forward?*.

DPIA evaluation

The DPIA produced for this case study will be evaluated in two steps. First, it will be evaluated as part of a series of staff training sessions (see Section 7.2.10) and second, it will be evaluated in a series of seminars to be held as part of the end of case study workshop planned to take place at the end of the case study (see Section 7.2.10).

Staff Training

As part of the GDPR implementation staff training will be provided (see Section 7.2.9).

The intention was that the staff training will serve a dual purpose;

1. The training sessions will be used to update staff on the changes brought in by GDPR, help them identify ways they can incorporate these changes into their working practices;
2. Training sessions will be used as an opportunity to evaluate the DPIA framework.

To pre-empt potential resistance to change (Demirci 2016) and help staff gain as much knowledge as possible from the training, the learning material for the staff training sessions will be designed to focus on how staff can be encouraged to embed good privacy practices into their working practices in data processing and handling. Adopting this type of approach to devise learning outcomes has been shown to have long lasting effectiveness (Moon 2004). In order to focus learning material to best support learning, students need to be empowered to believe themselves to be ”knowers”. This entails using material or examples from learners experience bank so that ”mutually constructive meaning” may be achieved from the training (Baxter Magolda 2004). To this end, transformative learning theory will be used to focus the learning material on events and scenarios observed by the researcher during site visits in the early phases of the project (Mezirow 1997). This, it is hoped will help staff relate better to the the training materials and help them see how they apply this to their work.

Workshop

A further evaluation will take place in a one day workshops. This workshop will be devised to inform other charities about GDPR and provide them with some hands-on experience of conducting DPIAs using the DPIA framework. The intention is that 50 attendees will be invited to attend the workshop from local charities within the local area. The workshop agenda can be found in Appendix D, Section D.7.

For the practical element of the workshop, attendees will be divided into four seminar groups to ensure each attendees gets an opportunity to participate in the discussion around privacy risks. Four facilitators will be recruited to facilitate these seminar sessions, the instructions for the facilitators will be attached in Appendix D, Section D.8.

As part of both the staff training session and the workshop seminar, participants will be given an information sheet (see Appendix C, Section D.6 for copy) and asked to complete a participant agreement form (see Section A.3). Further, as part of the session, they will be provided with a handout (this will be attached in Appendix D, Section D.9) and asked to provide feedback by completing an evaluation questionnaire, this will be attached in Section D.10.

After the session, facilitators will also be asked to provide feedback on how they felt the session went, a copy of the de-brief form will be attached in Appendix D, Section D.11.

7.3 DPIA Case Study

The case study was conducted over three months with the participants who took part in this study being made up of a mixture of management and staff. For the main GDPR implementation, 2 managers worked closely with the researcher in assessing the data holdings and conducting the initial evaluation of the framework. For the training sessions, all 29 staff and volunteers who work for the Charity took part over 3 staff training sessions. The case study culminated in a workshop for a group of 40 other local charities, where the results of the GDPR implementation were disseminated and a final evaluation of the DPIA Data Wheel took place (see Section 7.3.4).

The first step was to make a detailed GDPR implementation plan evaluating the Charity's readiness, and its ability to achieve data protection by design and default (DPbDD) and demonstrate GDPR compliance to the ICO.

7.3.1 DPIA Case Study: Phase 1 - Data Holdings Assessment

The first phase involved two parallel pieces of work, establishing what forms were used within the Charity to collect data and collecting staff stories. The aim of this phase was to establish what data is collected and how this data is collected, handled and processed within the Charity.

To establish what forms were used within the Charity, the project started with a meeting between the researcher and a member of the management team to find out a little bit more about where the Charity were with their GDPR implementation and establish what data was collected and processed within the Charity. This was a very informative meeting which formed the basis upon which the rest of the project was based. From this meeting it transpired that a draft data register had been started, detailing the types of data collected within the Charity. It also became evident that the majority of data collection and processing was paper based, securely stored in locked cabinets when not in use.

As part of this meeting, copies of the draft data register (4 different 'data collected' forms) and various templates and forms in use as part of the daily operations of the Charity (16 different templates) were provided to the researcher.

Staff User Stories

The next piece of work involved collecting staff stories. Storytelling as a research method involves collecting narratives or stories in order to understand people, their actions and ideas. Essentially, this involved getting staff to recount how they collect and handle data as part of their working day using user story methodology (see Chapter 3, Section 3.4.9). The term 'user stories' was changed to 'staff stories', as the Charity occasionally refer to their clients as 'users' (i.e. alcohol or drug users) and therefore, using this terminology caused some confusion.

To collect the staff stories, a spread sheet was created with 9 columns, one for each question (see Section 7.2.5 for a list of the questions). These spreadsheets, along with a participant information sheet (see Appendix C, Section C.5) and participant form (see Appendix A, Section A.3) asking staff to formally agree to take part in the research study, were sent to the Charity Manager, who circulated them to all staff and collated the responses. In total, 21 staff members, working in eleven different roles responded. The gender of the respondents was well balanced with approximately half of the respondents being male (9) and half female (10), two respondents chose not to provide gender details.

In the initial round of story collection,14 staff members took part with a further 7 responses received as part of the second round (described below). Once the staff stories had been collated, it became clear that, while some staff provided a lot of detail in their stories, others, were less forthcoming, giving one line answers. Further, from reviewing the responses, there was some confusion as to which forms and processes were referred to as staff used different terminology to describe what appeared to be the same documents or forms.

Therefore, a second round of questions was initiated to seek further elaboration and clarification on some of the terminology used. This took the form of returning the completed staff stories with an addition column containing questions, devised to seek confirmation of which forms and/or documents were being referred to and clarification on different aspects of the staff stories.

7.3.2 DPIA Case Study: Phase 2 - Analyse data holdings

Phase 2 sought to verify the data collected from phase 1. This involved analysing the data from the staff user stories and comparing this to the list of forms collated in the initial meeting with management. The analysis of the data provided in the staff stories showed that there were more template and forms in use within the Charity than those originally collected as part of the initial meeting. To address this, a second meeting was scheduled to update the list of forms and also seek further clarification on some of the terminology used as the forms were referred to in different ways by staff. At this second meeting, the research team was allowed access to the internal drive to download copies of all of the form templates used within the Charity. This resulted in a list of six folders (copied from the Charity intranet template folders) containing different templates and forms used as follows:

Access Admit Containing 5 *packs* of forms based on usage within the Charity. For example, Pack 2 contained a series of 19 different agreement and consent forms for all manner of agreements ranging from agreeing to conforming to the house no smoking rules to consent for sharing of data with family and/or other third parties;

Clients About Containing 44 different forms ranging from risk assessments (7 different types) to treatment plans;

Finance Containing 23 different forms relating to finance including expenses claim forms for staff and resident (client) signing over money for safekeeping to the house manager;

For Client Containing 14 forms relating to client's expectations, complaints procedure and relationships with staff and peers while in treatment;

House Containing 49 forms relating to house management including templates for various risk assessments; fire regulation records; creating house menus; staff rota's and various checklists;

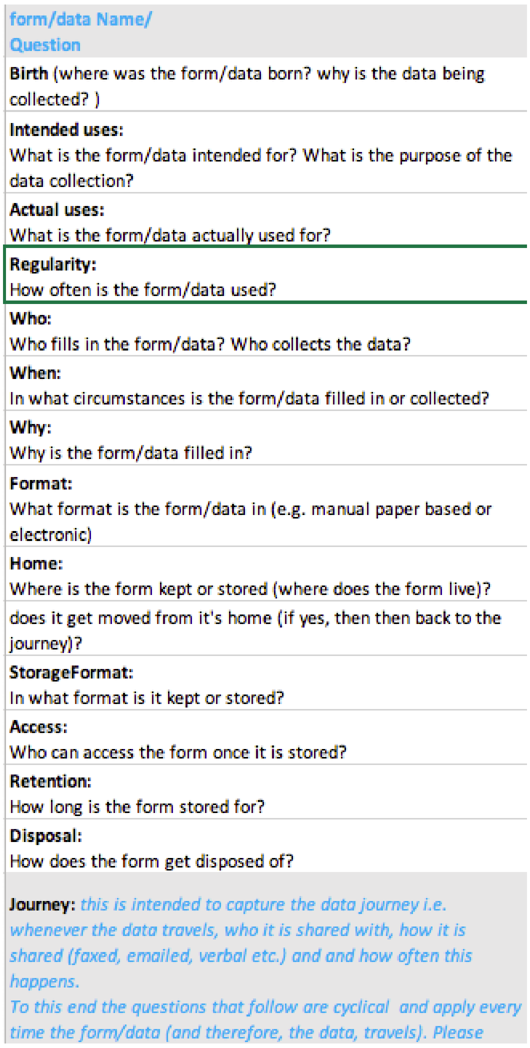
Managers Containing 21 forms relating to staff and client management including templates for various procedures such as disciplinary and grievance; audits; medication and volunteer and staff meetings.

In total there were 156 template and forms. However, many of these were duplicated across multiple folders. For example, the care plan consists of 23 different template forms, put together in packs according to the clients needs and/or their stage of treatment. Some of these 23 template forms therefore, may appear in multiple folders. Similarly, the client assessment consists of 12 separate assessments covering everything from client's personal details to addiction history and finance. Again, like the care plan, some of these assessment forms appeared in multiple folders.

In establishing what data holdings the Charity has and how data flows within the Charity, the first part of CS-RQ1 (see Section 7.2.2) was answered. To answer the second part, how closely these comply with GDPR standards, an assessment of how the data is transmitted was needed. This is what the second element of this phase was devised to establish.

Life of the Form

The second element in analysing the data holdings involved another spreadsheet, designed to capture the 'life of the form', i.e. the journey the data goes on as part of its lifecycle within the Charity. The life of the form spreadsheet captures details of the journey the data is likely to take during its lifetime with the organisation, allowing for up to 10 'journeys' (see Figure 7.1 for excerpt), devised to capture detailed insight into exactly how data travels within the Charity and which stakeholders have access to the data as part of those journeys.

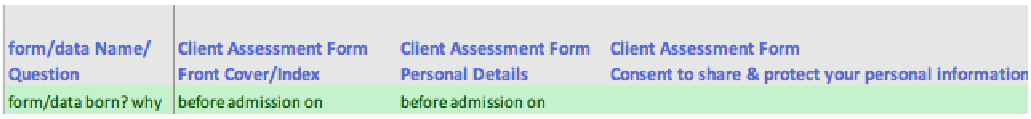


form/data Name/ Question
Birth (where was the form/data born? why is the data being collected?)
Intended uses: What is the form/data intended for? What is the purpose of the data collection?
Actual uses: What is the form/data actually used for?
Regularity: How often is the form/data used?
Who: Who fills in the form/data? Who collects the data?
When: In what circumstances is the form/data filled in or collected?
Why: Why is the form/data filled in?
Format: What format is the form/data in (e.g. manual paper based or electronic)
Home: Where is the form kept or stored (where does the form live)?
does it get moved from it's home (if yes, then then back to the journey)?
StorageFormat: In what format is it kept or stored?
Access: Who can access the form once it is stored?
Retention: How long is the form stored for?
Disposal: How does the form get disposed of?
Journey: *this is intended to capture the data journey i.e. whenever the data travels, who it is shared with, how it is shared (faxed, emailed, verbal etc.) and and how often this happens.* *To this end the questions that follow are cyclical and apply every time the form/data (and therefore, the data, travels). Please*

Figure 7.1: Excerpt - Life of the Form Spreadsheet

To make life a little easier for the two managers who had agreed to complete this exercise, it was decided to concentrate on just two of the template forms; (i) the care plan; and (ii) the client assessment. The reason for this decision was that, as a result of comparing the forms collected and the staff stories, it was evident that the main paperwork and documentation that may contain personal data that is handled regularly were the client assessment and care plan. Both these assessments are living documents that includes a lot of very detailed personal information, including full medical history (mental and physical) and details of the client's social, personal and cultural background as well as a list of professionals responsible or involved with their care both historically and currently. Both these documents form part of the contract between the client and the Charity.

Thus, a life of the form spreadsheet was pre-prepared for each of the two template forms (*care plan and client assessment*. On these a list of each of the forms was pre-completed across the top column in the first row so that each column would constitute a sub-form (i.e. a different aspect/data collection captured within the main assessment or care plan, a full list of these can be found in Appendix C, Section D.3). The reason for creating multiple columns was to allow for the fact that separate elements may go on different journeys. For example, a page or sub-form may be removed or shared for specific purposes such as faxing to external key professional staff who may be involved in the care of the client, could be captured as part of one of the journeys. The spread sheet allowed room for up to 10 journeys for each sub-form, see Figure 7.2 for example. These were then sent to the Managers for completion.



form/data Name/ Question	Client Assessment Form Front Cover/Index	Client Assessment Form Personal Details	Client Assessment Form Consent to share & protect your personal information
form/data born? why	before admission on	before admission on	

Figure 7.2: Client Assessment - pre-completed Life of the Form Spreadsheet

Once all the data was collated, an analysis was conducted, thereby confirming how data is handled within the Charity and thus, answering the second part of CS-RQ1. This revealed that records pertaining to the client's medication and the client register were the most frequently referred to documents. In addition, various methods of communication between staff and/or other stakeholders were mentioned as part of the staff stories. This information was compared to the completed life of the form spread sheets to provide a more detailed overview of the data, how it was used and the *journey* each form went on during its life cycle.The findings from the analysis highlighted a number of common processes and procedures undertaken by staff as part of their daily work and/or the journey the forms go on. These have been broken down into data relating to clients and data relating to staff and data handling processes and the findings for each group are outlined in Appendix D, Section D.4.

This analysis shows that there are a number of areas where the current processes result in potential risks to the data subject and/or data processes. For example, where a client file is removed from filing to travel, e.g. between houses, it was clear from the analysis that the file is transported but there was not sufficient detail in the stories or the journey to establish the exact details of how such travel might pose a risk, e.g. when the file is being transferred between houses, who is responsible and how securely is the file stored? To establish this, more detail would need to be elicited. Thus, this will need to be revisited as part of the DPIA risk assessment (see Section 7.3.4).

The analysis of the processes and procedures also showed that, while the Charity adhere to the current DPA, current processes would need revising and updating to be GDPR compliant. Thus, CS-P1 can be answered in the affirmative, the Charity's existing processes are not GDPR compliant (see Section 7.2.2).

Master Data Register

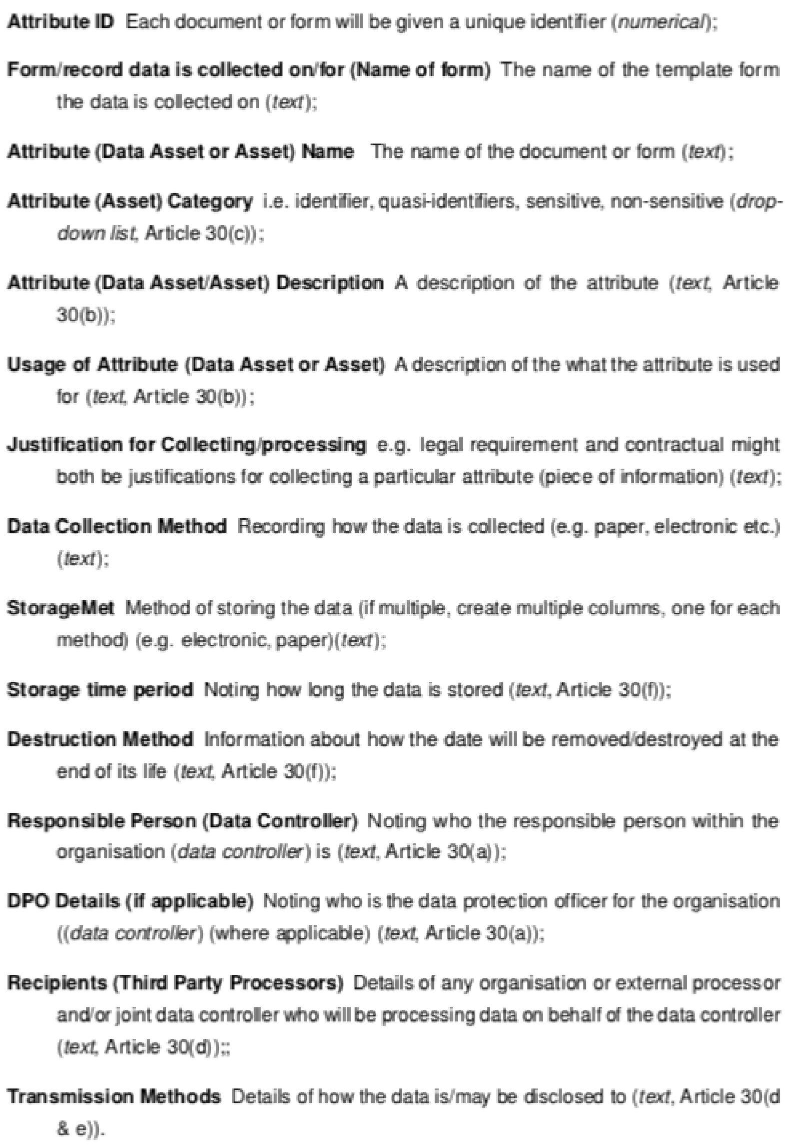
Although the Charity had made a start on creating a data register with their *data collected* forms, these were not fully compliant with GDPR as they did not contain enough information to meet the requirements of GDPR. For example, the headers on the data collected form only listed: the name of the data item, which form the data item was recorded on; what the form was used for; storage method and who the data item was disclosed to. GDPR Article 30: *"records of processing activities"* requires more details to be recorded such as details of any third-party stakeholders who might process the data (see Chapter 2, Section 2.4.4). Thus, CS-P2 holds true, the Charity's *current processes are inappropriate for supporting effective GDPR implementation.*

Therefore, as part of the analysis into the data holdings and pre-completing the life of the form spreadsheet for the managers, a draft data register was also created. This contained a list of all the attributes contained within each of the forms downloaded from the Charity. Each attribute was noted down separately initially and then later deduplicated. Further, as part of the draft, each attribute was categorised according to sensitivity. These categories were then revisited by the management team who reviewed the categories assigned. As part of this review the managers were asked to add a justification for why that attribute was collected by the Charity to the spreadsheet.

From this, a Master Data Register (MDR) was created. A description of the initial list of categories of data captured (columns) for the MDR spreadsheet can be found in Appendix C, Section 7.2.6. However, to ensure all obligatory data categories (GDPR Article 30) were captured, this original list was later changed and updated as part of the analysis. This involved changing, re-naming, removing and adding categories. The final MDR contained 16 categories (column headers in the spreadsheet). These are depicted in Figure 7.3.

The MDR aligns with the data register section of the DPIA so that all data collected as part of any new or existing DPIA can be copied and pasted straight into the MDR, thereby avoiding duplication of effort. The final MDR was handed over to the Charity. It contains details about each data attribute (individual data item) within each form used by the Charity, categorises according to data sensitivity. It also contains justification(s) for why each attribute is collected and details about how and when the data is collected, stored and destroyed and who is responsible for processing the data.

This final MDR contained 997 individual data items, categorised according to data sensitivity. Each attribute was justified to explain why that piece of data is collected. For example, some data is collected to facilitate clients' (*data subjects'* treatment needs, while other data is collected to meet a contractual or legal obligation such as to satisfy obligatory reporting requirement to the CQC (Care Quality Commission (CQC) 2018) or National Drug Treatment Monitoring System (National Drug Evidence Centre 2018). In creating this MDR, part of CS-RQ2 was answered (see Section 7.3), as the MDR provided the



Attribute ID Each document or form will be given a unique identifier (*numerical*);

Form/record data is collected on/for (Name of form) The name of the template form the data is collected on (*text*);

Attribute (Data Asset or Asset) Name The name of the document or form (*text*);

Attribute (Asset) Category i.e. identifier, quasi-identifiers, sensitive, non-sensitive (*drop-down list*, Article 30(c));

Attribute (Data Asset/Asset) Description A description of the attribute (*text*, Article 30(b));

Usage of Attribute (Data Asset or Asset) A description of the what the attribute is used for (*text*, Article 30(b));

Justification for Collecting/processing e.g. legal requirement and contractual might both be justifications for collecting a particular attribute (piece of information) (*text*);

Data Collection Method Recording how the data is collected (e.g. paper, electronic etc.) (*text*);

StorageMet Method of storing the data (if multiple, create multiple columns, one for each method) (e.g. electronic, paper)(*text*);

Storage time period Noting how long the data is stored (*text*, Article 30(f));

Destruction Method Information about how the date will be removed/destroyed at the end of its life (*text*, Article 30(f));

Responsible Person (Data Controller) Noting who the responsible person within the organisation (*data controller*) is (*text*, Article 30(a));

DPO Details (if applicable) Noting who is the data protection officer for the organisation ((*data controller*) (where applicable) (*text*, Article 30(a));

Recipients (Third Party Processors) Details of any organisation or external processor and/or joint data controller who will be processing data on behalf of the data controller (*text*, Article 30(d));;

Transmission Methods Details of how the data is/may be disclosed to (*text*, Article 30(d & e)).

Figure 7.3: Master Data Register Categories

Charity with the ability to demonstrate compliance with their obligation to keep “records of processing activities” (GDPR, Article 30).

7.3.3 DPIA Case Study: Phase 3 - GDPR Process Guidance

In this phase, existing processes were reviewed to determine whether and how these should be revised to ensure GDPR compliance was adhered to. As part of this a number of actions were taken:

1. The Charity’s Privacy Policies were reviewed and revised (see Section 7.3.3);
2. the process for obtaining consent from Clients was reviewed and revised (see Section 7.3.3);

3. a process for responding to requests from clients (*data subjects*) for access, erasure and data portability was created (see Section 7.3.3);
4. staff training on GDPR was arranged (see Sections 7.3.3 and 7.3.4);
5. the DPIA process created in phase 4 (see Section 7.3.4) was trialled and implemented.

Each of these are discussed in more detail in the following sections.

Privacy Policy Review

As part of reviewing existing processes, the privacy policy of the Charity was reviewed. While the existing privacy policy was in line with Data Protection Act 1998, it did not meet the requirement of expressing clearly, in plain language (see Chapter 2, Section 2.4.4), what data the Charity collect from clients and how this is used. Therefore, a new policy was devised to meet these requirements. This new policy is presented in plain language and includes a number of sections, a list of which can be found in Appendix D, Section D.2.

Consent and Lawful basis for processing

The issue of consent is a potential problem for the Charity in that the Charity works with vulnerable adults who may give consent to processing initially but later withdraw their consent or even claim consent was not freely given. This could be perceived as contrary to Article 7 of GDPR (see Chapter 2, Section 2.4.4). For example, a client may claim that they were not capable of giving informed consent at the time or they may claim they lacked sufficient mental capacity to freely give consent. Thus, there is potential that the Charity's clients (or someone else on their behalf) may argue that consent was not freely given, because there is an imbalance between the client (*"the data subject"*) and the Charity (as *"the data controller"*) who provide the client with treatment and thus, have a level of control over the clients and their actions while they are in treatment (Article 7, Recital 42).

To address this, a meeting was convened to understand exactly what consent is collected from clients (data subjects); how this is collected and used and what procedures are in place to allow clients to withdraw their consent. Based upon this meeting and careful study of the legislation, it was determined that the solution lies in looking at the legal basis for processing the data in the first place (see Section 2.4.4). For the Charity, they only process data in order to provide effective treatment to their clients. Although the Charity does also make sure they obtain informed consent from all their clients, this is not the main reason for processing the data. Prior to entering treatment, all clients complete a 'Client Assessment' and 'Care Plan' as part of the signing up for treatment process. Both of these documents subsequently form part of the contract between the client and the Charity. The data collected in these documents is required and necessary in order for the Charity to provide effective treatment, they cannot help the client unless they are aware of the full history to their addiction and the surrounding circumstances. Thus, based on this, the Charity can argue that the primary legal basis for processing these data is contractual (Article 6(1b)), they cannot perform their work without this information.

However, that does not mean that consent is not still required, as some aspects of sharing the data may not be required to perform the contract. For example, family members may wish to be kept informed of how the client is doing in treatment which is not a prerequisite requirement for providing treatment. Therefore, for those aspects, consent will still be required from the client. For this type of secondary sharing of the data, informed consent is a requirement under GDPR (Article 7(2)) which states:

"If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding."

To this end, the Charity, as part of the contract, obtain granular informed consent for who they may or may not divulge information to from the client. Therefore, the withdraw consent section was also amended to provide a granular 'withdraw consent' section, allowing clients to withdraw easily, together with the following statement added to top of the consent form:

"Whilst you are in treatment, we may at times need to discuss your progress with other professionals in order to ensure that you get the best service however, we will only do so with your consent. Please indicate which of the following professionals you would be willing for us to talk to, if necessary. You have the right to withdraw your consent at any time.

Withdrawing consent:
If you wish to withdraw your consent, please inform a member of staff and complete the 'withdraw consent' section indicating which professionals you no longer consent to us liaising with. If you choose to withdraw your consent the staff will explain to you how this may impact of the services the Charity can provide to you."

Further, it was agreed that the Charity would create a protocol for how any requests for access, erasure and data portability etc. by a client would be actioned. This will help the Charity in being able to demonstrate compliance and ensure that a thorough, repeatable procedure is in place to deal with a data subject (client) invoking one of their rights under GDPR Section 3 (*"Rectification and Erasure"*, see also Section 2.4.4). In addition, it was agreed that staff training should be provided to inform staff about GDPR, consent and the new protocols; what these mean for the organisation; and for them as staff, and as individuals.

Preparing for the Staff Training

As part of the GDPR implementation, it was agreed that staff would benefit from training on what GDPR would mean for them, their role, the processes they use currently and how best to make sure they adopted appropriate new methods where needed. To this end, a series of training sessions were arranged for staff to attend.

People in general do not like change and therefore, resistance to change can become a big problem for organisations when trying to introduce change processes or practices within the organisation (Demirci 2016). Conscious of this, the learning material for the training session was designed to be focused on the learning outcomes, how to embed good privacy practices into their every-day work when handling data. This approach has been shown to have maximum impact in the long term (Moon 2004) (see Appendix C, Section 7.2.10). Thus, transformative learning theory was used to devise the learning material so that it was centred around scenarios and events that had been observed during the earlier phases of the project, so that staff could relate the knowledge acquired (learning outcomes) to their day-to-day tasks (Mezirow 1997).

The training session itself was devised in two parts. The first part involved providing staff with an overview of GDPR, explaining how this is different from existing data protection regulations and what this will mean in terms of how they go about their daily activities and their work. Using the concept of *"communities of practice"* (i.e. groups of people who share a profession) to negotiate meaning and create knowledge can be very effective (Wenger 1998). Wenger contends that learning occurs where people engage in mutual relationships or engagement to gain and share knowledge. Thus, having spent some time investigating how staff currently handled data, this knowledge was used to devise material that would help staff to critically examine their current practices in a safe learning environment, thereby helping them recognise areas where they may not currently handle data correctly, so that they can adopt compliant practices going forward. This approach has been shown to encourage lifelong learning in adult learners (Christie et al. 2015), therefore, it is hoped that the learning gained might inspire staff to continue learning going forward.

7.3.4 DPIA Case Study: Phase 4 - DPIA

The obligation to conduct a data protection impact assessment (DPIA) is a new requirement brought in by GDPR. A number of differences exist between the two types of privacy assessment, the main one being the perspective from which the risk assessment should be conducted. The objective of conducting a PIA is to assess privacy risks from the perspective of the organisation whereas the DPIA asks that privacy risks are assessed from the perspective of the data subject (see Chapter 2, Section 7.2). Thus, CS-P3 can be answered in the affirmative as the Charity have not, so far conducted DPIAs as standard (see Section 7.2.2) and hence, why creating a DPIA has been included as part of this case study.

Thus, phase 4 sought to answer CS-RQ3 by creating a standardised DPIA process for assessing privacy risks that aligns with the meta-model (Chapter 5) and CLIFOD (Chapter 6).

Creating the DPIA Data Wheel

The creation of the DPIA Data Wheel started with an analysis looking at how CLIFOD could be adapted to incorporate the DPIA process that was made compulsory under GDPR. A DPIA is a privacy risk assessment aimed at assessing what privacy risks to the individual might be associated with a processing activity (see Chapter 7, Section 7.2). This is similar to some of the work already done on creating a privacy risk assessment framework for making decisions around open data (the meta-model of CI and the CLIFOD framework, see Chapters 5 and 6, however, it also differs in that the DPIA covers both existing and new data processing activities and how this is to be done has been defined as part of the GDPR regulation (Article 25).

DPIA Data Wheel - Design

This work stared by looking at how the requirements of a new system or dataset should be assessed according to the DPIA guidance provided by the ICO (ICO 2018a) and the wording of the GDPR regulation itself. This resulted in “the DPIA Data Wheel”, a DPIA framework based on previous work creating the meta-model created in Chapter 5 and the CLIFOD questionnaire developed in Chapter 6; GDPR; and the guidance provided by the ICO on how to conduct DPIAs (ICO 2018a), see Figure 7.4. Screen shots providing a visualisation of this process can be found in Appendix D, Section D.1.

Based on the initial conversion from GDPR to the DPIA Data Wheel depicted in Figure 7.4, the questionnaire itself was fleshed out to better guide the practitioner through how to conduct a DPIA. The DPIA Data Wheel incorporates questions about all the aspects of CI. For the context element, the questions devised for CLIFOD to elicit details about the prevailing and surrounding context were used.This framework was presented as a paper prototype (see Chapter 3, Section 3.4.11) in a spreadsheet.

The DPIA Data Wheel spreadsheet

The DPIA Data Wheel asks a series of questions relating to the data devised to provide a full overview of the system, process or project being assessed. The full DPIA Data Wheel

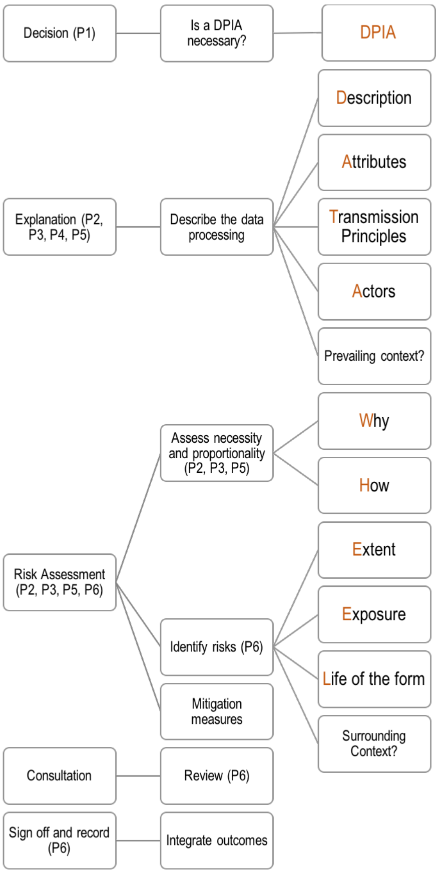


Figure 7.4: From GDPR & ICO to the DPIA Data Wheel

is presented in a spreadsheet consisting of 5 worksheets (*tabs*), the last containing the various drop-down lists within the spreadsheet. For information purposes, this has been left so practitioners can view this information. These are:

Worksheet 1: Need for a DPIA

This forms the starting point for conducting the DPIA. This is the starting point of the DPIA framework. This is a short assessment that will help the practitioner conducting the assessment to determine whether or not a DPIA is required for their system, project or process. The questions asked in this section can be found in Table 7.1.

In completing the pre-assessment *need for a DPIA* questions, if the outcome is a yes answer (i.e. a DPIA is compulsory or advisable), the practitioner will move to the second tab to begin the DPIA assessment and continuing through the rest of the framework. A NO answer, will end the process but ask the practitioner to provide an explanation for why a DPIA is not considered necessary and record this in the answer column.

In addition, a process model of the DPIA Data Wheel, depicted as a wheel, was also included on this tab to provide a visual representation of the DPIA framework. This can be found in Figure 7.5.

	Need for a DPIA	Question	Answer	(Explanation)
D	**Data**	Will a new system, project or process be implemented that involves collecting, processing, transmitting, sharing and/or storing of data or are there any changes to an existing system, project or process?	”yes/no” drop-down	
P	**Protection**	Looking at the ”when should we conduct a DPIA” list opposite, will the system, project or process involve processing or using any of the types of data listed in the first column	drop down (see Section D.9.2 for options available	(”DPIA Compulsory”)
I	**Impact**	Looking at the ”when should we conduct a DPIA” list opposite, will the system, project or process involve processing or using any of the types of data listed in the second column	drop down (see Section D.9.2 for options available	(”DPIA Advisable”)
A	**Assessment**	Based on the answers above, DPIA required?		

Table 7.1: DPIA Data Wheel - Is DPIA required

Worksheet 2: Data Wheel

Where a *yes* answer has been recorded and DPIA is required, the next step involves completing the DPIA Data Wheel. This is the privacy risk assessment for the process, system or project being assessed. The questions asked here can be found in Tables 7.2 and 7.3.

tab layout The spreadsheet is designed to have 13 columns, 5 of which are prepopulated as per Tables 7.2 and 7.3, followed by 8 columns for practitioner’s to complete. These are:

Answers providing a space for practitioners to answer the question;

Risks Identified please note each risk identified on a separate row. If necessary, add additional rows to the spreadsheet.;

Risk Likelihood i.e. likelihood of harm. This is a drop-down with the following selections: *Unlikely*; *Possible*; or *Highly Likely*;

Risk Score Severity : This is a drop-down with the following selections: *Minimal*; *Significant*; or *Severe*;

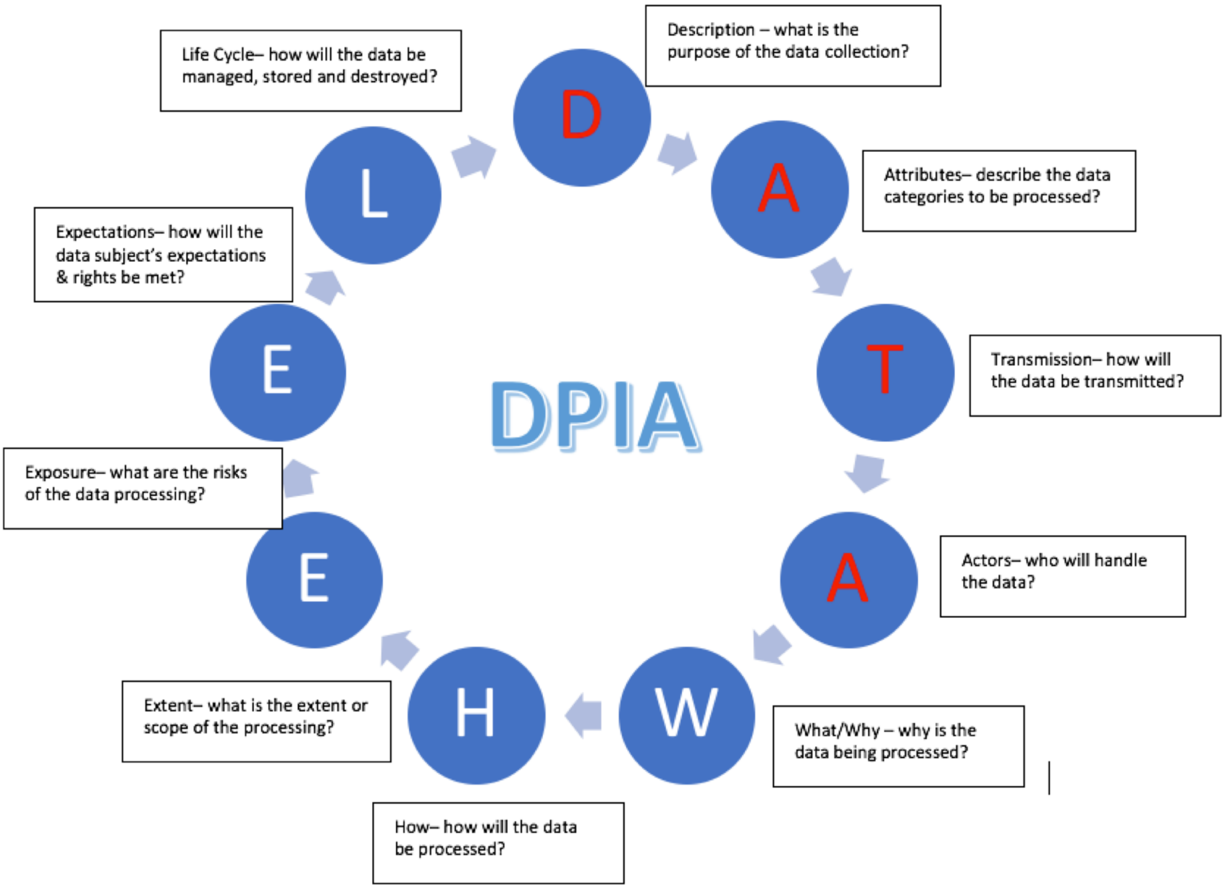

Description – what is the purpose of the data collection?
D
Attributes– describe the data categories to be processed?
A
Transmission– how will the data be transmitted?
T
Actors– who will handle the data?
A
What/Why – why is the data being processed?
W
How– how will the data be processed?
H
Extent– what is the extent or scope of the processing?
E
Exposure– what are the risks of the data processing?
E
Expectations– how will the data subject's expectations & rights be met?
L
Life Cycle– how will the data be managed, stored and destroyed?
DPIA

Figure 7.5: The DPIA Data Wheel

Data Subject Risk - Impact of risk on the data subject : This is a drop-down with the following selections: *Minimal*; *Significant*; or *Severe*;

Organisational Risk - Impact of risk on the organisation : This is a drop-down with the following selections: *Minimal*; *Significant*; or *Severe*;

Overall Risk Score : Taking all the scores into account for each risk - please rate the risk overall. This is a drop-down with the following selections: *High*; *Medium*; or *Low*;

Mitigation Please note any mitigation strategy that might be applied to reduce or eliminate the risk.

Worksheet 3: Data Register

The data register forms part of the answer to question 4 where practitioners are asked to complete the data register for any new system or process being assessed. The data register was derived from the Master Data Register (MDR) created as part of Phase 1 (see Chapter 7, Section 7.3.2). The idea behind this is that, once complete, the practitioner

Question No.	CLIFOD Phase		DPIA Data Wheel	Questions
[Q1]	**Explanation**	**D**	**Description**	What is the purpose of the data collection/processing?
[Q2]				What is the system, process, project or dataset about (subject matter/context)?
[Q3]		**A**	**Attributes**	Please describe what data will/has be/been collected?
[Q4]				If this is a new project or system, please complete the Data Register tab with details of each attribute
[Q5]		**T**	**Transmission Principles**	How will the data be processed internally within the organisation?
[Q6]				Please describe the information flows with internal stakeholders (then please go to life of the form tab and complete this to gather more detailed information about the data collection/project/system being assessed
[Q7]				Will the data be transmitted to external stakeholders?
[Q8]				Please describe the information flows with external stakeholders
[Q9]		**A**	**Actors - internal**	Please provide a list of all staff/roles who will be processing/using the data - you may wish to add extra rows for each additional data processor(s) - (Data Processor(s))
[Q10]				Responsible Person - name and contact details of the designated Data Controller
[Q11]				Data Owner (if different from Processor)
[Q12]				Data Subject(s)
[Q13]				Name & contact details of Data Protection Officer (DPO) if applicable
[Q14]				Other (please explain role and relationship)
[Q15]			**Actors - external**	Please list of all/any third parties with whom the data will be shared - you may wish to add extra rows for each additional data processor(s) - (External Data Processor(s))
Q16]				External Responsible Person - name and contact details of the designated Data Controller
[Q17]				External Data Owner (if different from Processor)
[Q18]				External Data Subject(s)
[Q20]				Name & contact details of third-party Data Protection Officer (DPO) if applicable
[Q21]				External Other (please explain role and relationship)
[Q22]		**Context**	**Prevailing Context**	What are the relationships between actors? (please describe role and relationships between data- controller, processor(s) and Subjects)
[Q23]				What was/is the social context of the data collection e.g. school would be educational context, council tax would be for tax collection etc.
[Q24]				Is/was data collected directly from data subject
[Q25]				Are there any overriding considerations as to why processing should be allowed despite lack of consent/limited consent?

Table 7.2: DATA

will be able to ”lift” the information directly from the data register into the organisation’s MDR. Further, the information gathered on the data register is intended to be used to help inform the risk assessment in the data wheel.

Worksheet 4: Life of the form

The Life of the Form has been derived from the life of the form devised as part of the GDPR implementation (see Chapter 7, Section 7.3.2). Question 6 in the Data Wheel asks that practitioners complete this tab. The intention is that the information gathered here will help inform the risk assessment by encouraging practitioners to think about how the data travels within their organisation. (*the “transmission principles”*). It is through considering the ‘journey’ the data might take within the system, project or process during its lifetime that practitioners can enquire more deeply into the data flows and thereby, glean valuable insight into where there may be potential privacy risks.

Question No.	CLIFOD Phase		DPIA Data Wheel	Questions
[Q26]	**Risk Assessment**	**W**	**What/Why**	What are the benefits of the processing for the data subject?
[Q27]				What is/are the desired effects of the processing for the data subject?
[Q28]				What is the desired effect of the processing for the organisation?
[Q29]				What is the primary legal basis for collecting/processing or handling the data?
[Q30]				If applicable, what is the secondary legal basis for collecting/processing or handling the data?
[Q31]				What are/is the desired effects of the processing for the organisation?
[Q32]				Why is the data being collected/processed? (purpose limitation/relevance)
[Q33]				How much data will be collected?
[Q34]				Who is responsible for security around manual data handling, processing or storing?
[Q35]				How will the data be collected?
[Q36]				How is/will the data be accessed and used/processed?
[Q37]		**E**	**Extent**	What is the extent of the processing - will we require consent?
[Q38]				Identify any risks of the dataset or individual attributes, being obtained or accessed by unauthorised parties or means in such a way that they can pose a risk a data subject? (please be specific as to what risk and how this might pose a new risk)
[Q39]				Security - how is/will the data be protected and kept safe?
[Q40]		**E**	**Exposure**	Who is responsible for access control and data security (electronic data handling, processing or storing)?
[Q41]				What are the risks of the dataset or attributes, being linked to external data in such a way that they can pose a risk of contributing to re-identification of a data subject? (please be specific as to what risk and how this might pose a new risk)
[Q42]				How is/will compliance be evidenced?
[Q43]				How will we meet stakeholders expectations and adhere to data subjects rights?
[Q44]		**L**	**Life Cycle**	What is the expected data life cycle? i.e. for how long will the data be *live* and processed
[Q45]				Has data limitation been considered, is all the data being collected necessary?
[Q46]				How is/will the data be stored?
[Q47]				Life of the form - what journey(s) will the data take as part of its lifecycle?
[Q48]				How do/will you delete the data?
[Q49]				How long will the data be kept?
		Context	**Surrounding Context**	Could the data collection or processing be perceived to infringe upon:
[Q50]				any political values? (please also explain whose values and how they might be infringed upon)
[Q51]				any social values? (please explain what those values are, whose values they are and how these might be infringed upon)
[Q52]				any moral values? (please explain what those moral values are, whose moral values and how these might be infringed upon)
[Q53]				any legal compliance? (if yes, please clarify what legal compliance might be breached, how and who would be affected)
[Q54]				any belief systems that could adversely be affected by processing? (please explain what belief system could be affected, how and who would might be adversely affected)
[Q55]				What are the positive values that data processing will bring/enhance. These may include commercial gain, improved transparency, meeting legal obligation etc.
[Q56]				impose any form of power imbalance? (please explain how such a power balance might arise, who might be affected and how these might be imposed)
[Q57]				pose a threat to the autonomy or freedom of the data subject(s)? (please explain how this might pose a threat, who could be affected and how)
[Q58]				result in any form of discrimination (please explain what form of discrimination, how this might occur and who would might be adversely affected)
[Q59]				Grant or afford any privileges or prerogatives to the referer, the data- processor(s), controller(s) or originator(s) that may benefit or be perceived to benefit those parties as a result of processing, sharing or publishing the data?
[Q60]				Are there any overriding legal, moral or ethical reasons why processing should be allowed even if there is a risk of re-identification?
[Q61]				Is there a reasonable expectation on the part of the data subject(s) and/or data controller(s) of data being kept confidential and not shared?

Table 7.3: WHEEL

Worksheet 5: List

This contains list of all the drop-down menus that form part of the assessment on the other tabs. This includes:

- Yes/no;
- DPIA Required where (see Section D.9.2 for list);
- DPIA Advisable where (see Section D.9.2 for list);
- Lawfulness of processing (see Section 2.4.4 for list);
- Attribute types: *PI* personal identifiers; *QI* Quasi identifiers; *S* Sensitive; *NS* Non-sensitive (see Section 5.4.1); or *C* Combination of all;
- Risk Score Likelihood (see above for list);
- Risk Score Severity/impact: (see above for list);
- Overall Risk Score: (see above for list);
- Attribute classifications: *PI* personal identifiers; *QI* Quasi identifiers; *S* Sensitive; *NS* Non-sensitive; or *O* Other (If unsure or free text field, use 'other' (S) to indicate further manual review required).

Thus, the final DPIA Data Wheel is a step-by-step assessment that takes practitioners through the process of conducting a DPIA (see Figure 7.6).

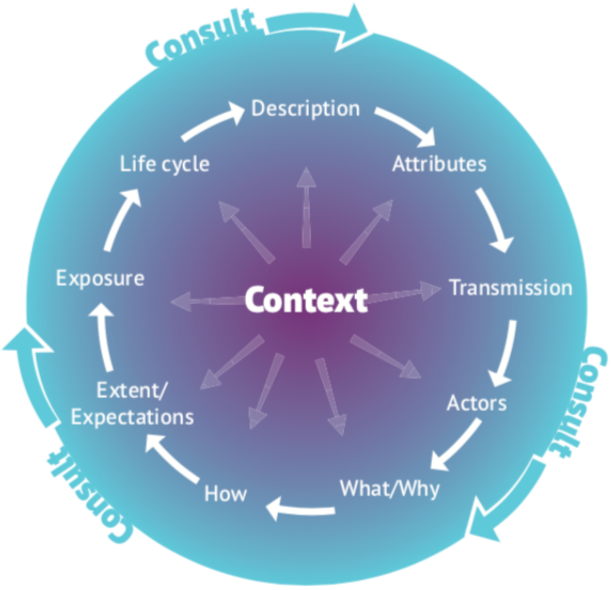


Figure 7.6: DPIA Data Wheel

Within the DPIA Data Wheel, incorporating the data register and the life of the form questionnaires, used in phases 1 and 2, facilitates the gathering of comprehensive background information about the data, the actors and the transmission principles to inform the risk assessment in the Data Wheel. For the consultation element, this has been 'removed' from the process and instead used as a *loop* that circumscribe the DPIA Data Wheel to show that this element is a continuous and ongoing process. The idea behind consultation is to first, ensure that all relevant stakeholders who have an input or interest into the process, system or project have an opportunity to help identify potential privacy risks and be consulted about such risks so that they too can take part in devising the mitigation strategies. This also addresses the finding from Chapter 6 that multiple practitioners should be involved in the privacy assessment (see Section 6.6.4). In the case study, this element was completed through the workshop and staff training sessions (see Section 7.3.4).

Evaluating the DPIA Data Wheel

The protocol stipulated that the DPIA framework should be evaluated through staff training and workshop sessions (see Section 7.2.10). However, during the case study the design and format of the DPIA Data Wheel evolved throughout the project and therefore, arguably, the evaluation began with the first iteration of the questionnaire (see Sections 7.3.4 and 7.3.4). Therefore, all of the evaluation steps taken are described in this section.

Evaluation 1

The first evaluation consisted of sending the DPIA Data Wheel spreadsheet to:

Practitioners here three practitioners with a background in data management and/or security were approached and asked to comment on the layout, format and questions of the spreadsheet;

Academic in parallel with this, three Academics were also approached for comment;

Peers finally, three peers (fellow PhD students with expertise in risk and/or security) were approached for comment.

As a result of feedback from these evaluators, some questions were changed, some questions were rephrased slightly and an additional worksheet was added at the beginning of the spreadsheet, named *Instructions and Definitions*. This worksheet contained instructions for how to work through the spreadsheet and what each worksheet covers. Further, there is a definition section with explanations for the terminology used within the framework. These instructions and the definitions can be found in Appendix D, Section D.14.

Evaluation 2

The second evaluation took part as part of completing the Master Data Register (MDR)). As part of the evaluation of the DPIA Framework, a DPIA Data Wheel spread sheet was created for the *Care Plan* and the *Client Assessment*. On each of these pre-completed DPIAs, the Data Register and the Life of the Form worksheets were pre-populated with the information provided as part of phase 1 and 2, i.e. the list of attributes collated from the staff stories and the forms provided that the Charity use and the completed life of the form answers that the Chief Financial Officer (CFO) and House Manager had provided. A meeting was then arranged to go through these two DPIAs to complete the DPIA and Data Wheel sections of the spreadsheet.

The CFO and House Manager completed the Data Wheel for the care plan and the client assessment forms and, in doing so, identified and scored a number of potential risks before returning the completed spreadsheet to the researcher. As part of completing the DPIA Data Wheel, the CFO and House Manager identified 21 risks for the care plan and 25 risks for the client assessment and noted potential mitigations for each of these on the DPIA. These had been recorded on the spreadsheet next to the answers to the questions, one risk for each answer. This was of course, not the intention, the idea is for all risks to be identified based on the answers, not just one risk per line or row. However, the Managers thought the layout suggested that privacy risks needed to be identified to align with each answer only rather than as a whole. However, in having the risks on the same worksheet as the Data Wheel, the Managers did not think the link between the risk assessment and the information from the life of the form and data register was evident. Thus, as these were meant to be used to inform the privacy risk assessment, this section did not work as intended resulting in separating the privacy risk section out for better overview. Therefore, in light of the CFO and the House Manager having misunderstood how to complete the risks section, the presentation was altered, moving the privacy risks into a separate worksheet to avoid confusion.

Another amendment to this original design was made on the data register where 3 additional columns were added for inputting justifications, as one was not always sufficient (see Section 7.3.2), there can be more than one reason for why a particular attribute is collected and the Charity wanted to capture this to strengthen their case for justification. In addition, based on feedback from the Managers, some further changes were made to the DPIA framework. Some of the questions were reworded slightly in the Data Wheel, one question was removed as it was, effectively, a duplicate, and another added. The final questions of the Data Wheel can be found in Appendix D, Section D.5.

The amended format of the Data Wheel worksheet on the spreadsheet therefore included the following worksheets:

Instructions & Definitions ;

Worksheet 1 - Need for a DPIA ;

Worksheet 2 - DPIA Data Wheel as in the original design (see Tables 7.2 and 7.3), on this worksheet the first 5 pre-populated columns appeared together with a column for providing answers to the questions and the following additional columns were added to the Data Wheel worksheet on the spreadsheet:

- DPIA Review (Answers);
- Reviewed by;
- Review Date.

Worksheet 3 - Data Register ;

Worksheet 4 - Life of the Form ;

Worksheet 5 - Privacy Risks this worksheet consists of the following column headers:

- Risks Identified: please note each risk identified on a separate row. If necessary, add additional rows to the spreadsheet;
- Risk Likelihood (Likelihood of harm): drop-down selection *unlikely, possible or highly likely*;
- Risk Severity (Severity of harm): drop-down selection *minimal, significant or severe*;
- Data Subject Risk (Impact on data subject): drop-down selection *minimal, significant or severe*;
- Organisational Risk (Impact on organisation): drop-down selection *minimal, significant or severe*;
- Overall Risk Score, taking all the scores into account for each risk - please rate the risk overall: drop-down selection *high, medium, low*;
- Please note any mitigation strategies that might be applied to reduce or eliminate the risk;
- Action taken;
- Actioned by;
- Date.

List containing the list of drop down selections utilised in the various worksheets across the spreadsheet for information.

Evaluation 3 - Staff Training

The third evaluation formed part of the staff training sessions. The intention was that the staff training would serve three purposes: (i) to inform and encourage staff to adopt the changes required in processes and protocols required for successful incorporation of GDPR into the Charity; (ii) to further evaluate the DPIA framework; and (iii) ensuring that the final piece of the DPIA framework, to consult, could be met.

The training sessions were arranged as three separate sessions, one for the staff from each of the three houses that the Charity run. In total, 29 members of staff participated in the training sessions across all three houses. The session began with an introduction to GDPR that provided an overview of the regulation and what this might mean for staff as part of their daily work. As part of this staff were provided with a handout which provided an overview of GDPR and the DPIA (see Appendix D, Section D.9).

Following on from the introduction to GDPR, the second part of the sessions consisted of a group discussion about privacy risks, thereby using the staff training session as an opportunity to consult on the risks identified in the second evaluation and to elaborate on these through a discussion with staff about privacy risks. Because of time restrictions for the sessions, only one of the risk assessments was reviewed, the risks for the care plan. To this end, part of the handout consisted of the completed data wheel with the answers provided by management in evaluation 2, pre-completed in the answers column (see Tables D.2 and D.3 in Appendix D, Section D.9). This discussion centred on what risks staff perceived were in respect of the introduction of GDPR, in light of their daily practices, as part of their work and what occurred. This resulted in a list of 35 additional risks being elicited and potential mitigations for dealing with each of these being identified. These were all different to the list of risks identified in Evaluation 2 and therefore, these were added to the list of risk initially identified by the CFO and added to the risk register.

After the session, staff were asked to complete an evaluation survey to gauge how effective the sessions had been and whether the staff had found the learning useful. First, staff were asked *overall do you feel your knowledge has improved regarding GDPR?*, all the staff answered yes. The second question staff were asked was: *overall do you feel your knowledge has improved regarding DPIAs?*, here 93% responded yes, they felt that their knowledge had improved as a result of the training session. The survey then asked a series of questions about their experience of the training session, including asking how likely staff felt they were to integrate the knowledge gained into their daily work practices, most staff (80%) responded ’extremely’ or ’very likely’ (see figure 7.7), indicating that the training had been well received and effective.

When asked how much of the information conveyed regarding GDPR was new to them, 75% said half or more of the information was new to them. Similarly, for DPIAs, 72% felt the information regarding DPIAs was new to them. Finally, when staff were asked how

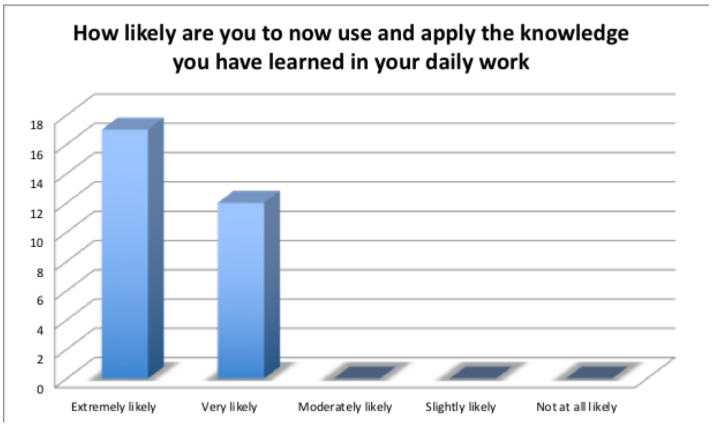


Figure 7.7: Likelihood of applying GDPR in workplace

comfortable they would feel about explaining DPIAs to a colleague or friend, only 62% felt comfortable doing so in comparison to 72% being comfortable in explaining GDPR to a colleague or friend.

Staff were also asked about what they liked about the training session and what they felt went well. To this, 10 staff felt that the session was very informative, and 14 staff specifically liked and commented that the group style and interactive nature was their main positive takeaway. What was disliked about the training session was that some felt there was too much information to take in (5), that the subject was complicated (2) and that there was not enough time allocated to the sessions (2). Appendix D, Section D.10 contains a copy of the evaluation form, and Section D.12 includes graphs for staff responses.

Finally, staff were asked if they had any further comments or suggestions. In this section, some staff used this as an opportunity to suggest *''regular updates''* or; *"maybe a regroup training session in a couple of months''*, while others commented; *''I will be mindful and start to prompt colleagues around having data around"*. Finally, one staff member suggested that: *"a company email to bullet point main points to keep it fresh in peoples' minds and to offer open dialogue to discuss new rules"* as a potential way to keep the momentum going around GDPR and DPIAs and making sure staff remain vigilant in their practices.

Evaluation 4 - DPIA Workshop

The fourth and final evaluation took place as part of a one-day event, a workshop where 50 local charities were invited to attend the one day workshop designed to to disseminate the findings from the GDPR implementation, consult with the wider stakeholder community (the Charity's peers) and further evaluate the DPIA Data Wheel. The agenda for the day can be found in Appendix D, Section D.7.

On the day of the workshop, there were 40 attendees from 11 different local charities participated in the workshop. Most of the attendees were charity workers with no expertise in privacy or security, with only 2 attendees claiming none of the information provided on GDPR was new to them and 85% reporting that more than half the information provided on DPIAs was new to them (based on evaluation feedback, see Appendix D, Section D.13).

Upon arrival, attendees were provided with a welcome pack that contained supporting notes and materials about GDPR and DPIAs as talked about in the workshop sessions (see Figure 7.8 for handout contents) together with a notepad, pen and mouse mat depicting the DPIA Data Wheel (see Figure 7.6).

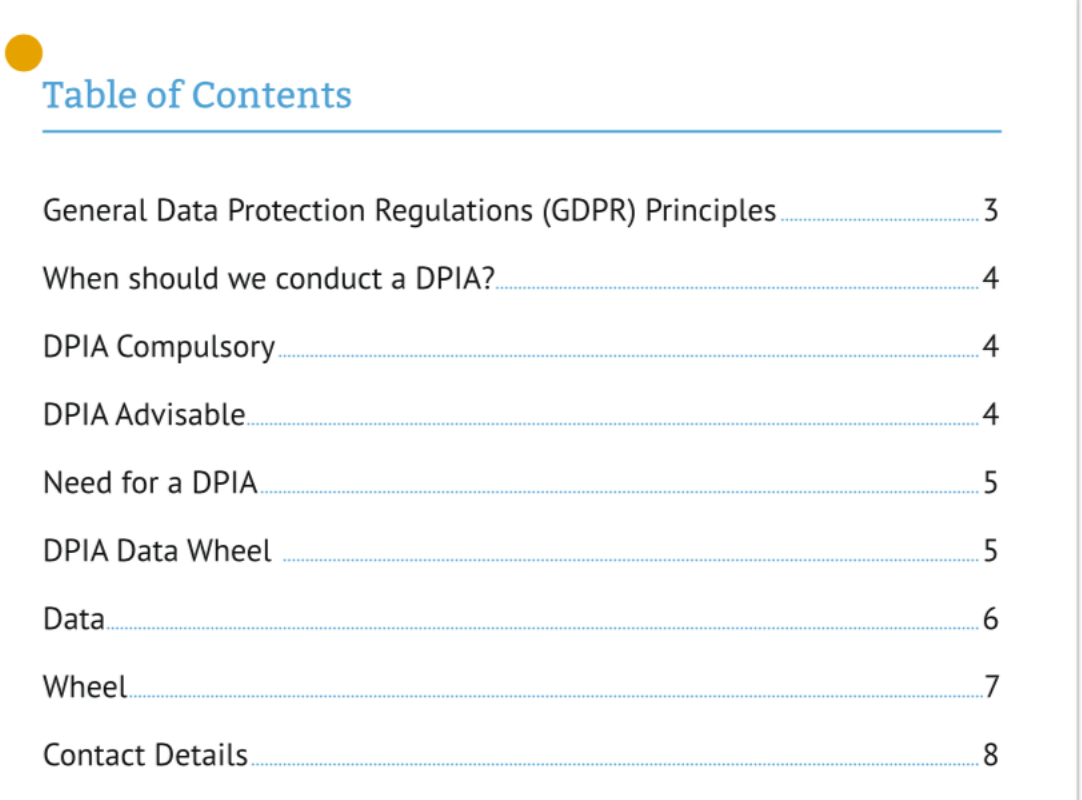


Table of Contents

General Data Protection Regulations (GDPR) Principles 3

When should we conduct a DPIA? 4

DPIA Compulsory 4

DPIA Advisable 4

Need for a DPIA 5

DPIA Data Wheel 5

Data 6

Wheel 7

Contact Details 8

Figure 7.8: Table of Contents Workshop Handout

The day started with a talk on GDPR and how the new regulation will affect charities This was followed by an overview of data protection impact assessments (DPIAs) and a step-by-step overview of the DPIA Data Wheel, after which attendees were divided into four seminar groups. For these seminar groups, the main researcher, assisted by three volunteers with backgrounds in security were recruited from the University, facilitated the sessions. The instructions given to the facilitators can be found in Appendix D, Section D.8. Each seminar group was provided with a handout consisting of a pre-completed DPIA Data Wheel, like the one given to staff in the training sessions (Appendix D, Section D.9) and informed that the Charity were looking to change the data processing format of the care plan from manual to electronic. The facilitators then went on to explain that their task, as a group, was to:

1. Complete the review section on the Data Wheel, re-assessing the answers provided

in the handout and discuss how these answers might change in light of the change in data processing format; and

2. Complete the privacy risk and discuss what privacy risks the change in data flow would be likely to bring about.

Re-assessment of Data Wheel

For most of the questions on the Data Wheel, one of the groups had determined that the change in data flow did not alter the answers already provided and therefore moved straight to the risk discussion. The other groups however, had identified a number of areas where there were some deviations from the original answers provided as follows:

Describe attributes for the personal identifiers one group had noted that *"Data will be the same but the quantity will be condensed as each item will only be recorded once e.g. name rather than multiple times on many forms"* (G1);

Transmission Principles these were highlighted as changing as a result of the change from manual to electronic processing, one group elaborating to comment that: *"the internal data flow would change as data will be collected electronically on i-pads"* (G2). Another group noted that transmission with external stakeholders would also be likely to change as files may now be shared and submitted electronically between stakeholders;

Actors one group highlighted that, while the actors would be the same, the way they communicate and share the data would likely change (see transmission principles);

Exposure One group identified that firstly, staff were responsible for access controls and data security and secondly, to set boundaries around data handling *"password protection of documents"* (G3) should be implemented;

Expectations here it had been noted that: *"electronic data will make it easier to deal with right of access/erasure as all the data will be electronic in one place"* (G1);

Life cycle data storage was noted as a change with data being stored electronically rather than in paper format, one group noting *"the security of the network where data is held electronically will change - data risks much higher once data is electronic so appropriate security will need to be put in place"* (G1).

Interestingly, no changes had been recorded in the context sections for any of the groups. This could be because of lack of understanding on the part of the participants and/or facilitators. This was highlighted by one of the facilitators as part of their feedback where they stated: *"I don't think the instructions were lacking it was more my lack of* *experience with facilitating that may have impeded the process"* (F2). However, the facilitators had been specifically instructed not to 'direct or lead' the discussion, rather, let the participants discuss the answers among themselves. This, in hindsight, might have been a mistake as a more in-depth discussion of context could have provided some insight and understanding of how context affects data flow, thereby helping participants appreciate the wider implications of assessing privacy risks. A similar sentiment was provided in the evaluation feedback provided by participants (see Appendix D, Section D.10), with one attendee commenting: *"Some areas seem grey and it is so individual to each charity it seems a minefield"* (A17).

Discussion of Privacy Risk

The second part of the seminar sessions proved much more successful. Here all the groups had elicited between 5 and 14 different privacy risks and identified mitigation strategies for how to minimise or avoid these. The risks identified ranged from *"hacking, theft, loss or fire"* (G2 & G1) to *"conflicting information between manual and electronic records during infrastructure transition period"* (G4) and *"data being processed and stored by third-party outside of the Charity's company access controls"* (G3). In total between the four groups, 35 different privacy risks were identified and scored. These risks were different to the risks identified as part of evaluations 2 and 3, which shows how different peer groups perceive the same situation from different perspectives. Thus, this illustrates the benefits and importance of wide consultation in that, the wider the consultation, the more detailed and in-depth the resulting privacy risk register will be.

The privacy risks identified as part of the seminar groups were later consolidate and, after the sessions, all the risks from across the evaluations, were amalgamated into one spreadsheet that was passed to the Charity for entering onto their privacy risk register. What was interesting was that all the risks elicited related to the organisation and how they should secure and safeguard the data rather than to the data subjects which is what the DPIA is supposed to elicit. This may be due to not enough emphasis being placed on this distinction in the training and instructions provided as part of the evaluations or, it may be because, practitioners are conversant with risks from the organisational perspective and therefore, more comfortable in this domain. Discussing risks from the perspective of the individual (data subject) is not something that has previously been considered. Thus, future work will need to look at how this aspect can be addressed to highlight this distinction and ensure practitioners understand the different perspectives better.

Workshop Feedback

Based on the feedback provided by the workshop attendees, most appeared to really enjoy the day although only half completed the evaluation forms. Of those that did respond, 75%

stated that they were very or extremely likely to recommend the session to colleagues or friends. The statistics on the responses from participants can be found in Appendix D, Section D.13. In the comments section one attendee commented the day provided: *"Clear, concise info & resources & ability to ask questions"* (A8), while another stated they would have liked more *"practical application and spreadsheet scenario's"* (A11). At the other end of the spectrum, one attendee wrote: *"Feel lost with it. As a small charity with low capacity but high data collection"* (A18).

However, in respect of the DPIA Data Wheel itself, attendees were very positive, one commenting: *"Thank you, we had found lack of guidance for GDPR for charities very frustrating so greatly appreciated this event"* (A20), indicating the format of the DPIA Data Wheel is an effective tool for conducting DPIA for non-specialist practitioners.

Similarly, the facilitators for the seminar sessions provided positive feedback on the Data Wheel itself, one stating: *"the Data Wheel is really helpful, probably for supporting security risk assessments too, and (within reason) you don't have to be too technical to follow the process"* (F1). Looking at the feedback in more detail, everyone felt the privacy risk discussion went well, with one commenting their favourite part was *"the open discussions within the group"* (F2) and another stating: *"the general discussion that was provoked from attendees due to framework was good, I think it made them adopt a different approach to eliciting potential risks. A much more methodical approach due to referencing the DPIA Data Wheel"* (F3). However, it was also commented that more time should have been devoted to the seminar sessions and to explaining the concepts in a bit more depth prior to the sessions, as highlighted by one of the facilitators who commented *"it was noticeable that industry attendees see a risk as a concern of something happening. In the session I did explain the components of risk (Asset + Threat x Vulnerability = The Risk), but most are not in-tune with the process of risk. ...this in part slowed down the session because once people were comfortable to share ideas and experiences, there were quite a few ideas of concerns/risks, but not much substance behind how it was actually a risk"* (F1). Arguably this level of detail is beyond what the DPIA Data Wheel is trying to elicit. The practitioners who are likely going to complete the assessment will not be experts in IT, privacy or security and therefore, will be unlikely to have the expertise to understand or discuss risk in such terms. However, it is possible that the fact that neither the idea of what a risk is, nor the context had not been discussed or explained in detail as part of the introduction to DPIA and GDPR could have influenced the outcomes and lack of discussion around context. Thus, if the workshop was to be repeated, this would be one aspect that should be reviewed as revised.

7.4 *DPIA Case Study: Summary of Findings*

This chapter describes how GDPR was implemented within a charitable organisation (*the Charity*). For this case study, three sub-questions were devised to inform and support the implementation: CS-RQ1-3, supported by three hypothesis: CS-P1-3 (see Section 7.2.2. The first question (CS-RQ1) was answered by gathering staff stories in phase one (see Section 7.3.1)) and analysing these in phase 2 (Section 7.3.2). This provided an overview of the Charity's data holdings and the data flows within the Charity. This analysis confirmed that, while existing processes for data handling were DPA compliant, they needed updating to meet GDPR standards, thus confirming the CS-P1 hypothesis that the Charity's existing processes were not GDPR compliant.

To address this the Charity's existing processes and privacy policies were reviewed and revised to establish what processes the Charity needed to implement to demonstrate GDPR compliance. This analysis confirmed CS-P2 to be true, existing processes were insufficient to support an effective GDPR implementation. To rectify this and answer CS-RQ2, a Master Data Register (MDA) was created in Section 7.3.2 and existing processes were reviewed and revised in phase 3 (see Section 7.3.3). This resulted in the creation of an updated privacy policy for both clients (*data subjects*) and staff that meet GDPR requirements (see Section 7.3.3) and a revised procedure for obtaining and recording consent (see Section 7.3.3).

As part of the implementation, three paper prototype (see Chapter 3, Section 3.4.11) spreadsheets were created: (i) *the staff user stories*, created as part of phase 1 (Section 7.3.1); *the Master Data Register* (Section 7.3.2); and *the life of the form* (Section 7.3.2). Of these, the latter two were incorporated into the DPIA Data Wheel, the data protection impact assessment framework created in phase 4 (see Section 7.3.4), to address the fact that the Charity did not conduct DPIAs as standard (confirming hypothesis 3 (CS-P3)). Therefore, a standardised DPIA was devised to provide the Charity with a repeatable, consistent DPIA process that they can use going forward. The resulting DPIA framework, the DPIA Data Wheel, provided the Charity with an empirically evaluated DPIA process, thereby enabling them to demonstrate compliance with GDPR Article 35 (see Chapter 2, Section 7.2) and answering CS-RQ3 (see Section 7.2.2).

This DPIA framework, was the main output from this case study, providing a comprehensive DPIA framework for assessing privacy risks to the data subject of a new or existing process, project or system (Figure 7.6). The DPIA Data Wheel was empirically evaluated through a series of trials carried out with the Charity management, staff and other charity sector peers (see Section 7.3.4). In total, across all the evaluations 25 risks were elicited for the client assessment and 91 risks were identified for the care plan, 3 of which were duplicates, resulting in 88 different risks being identified and mitigation strategies suggested for each of these. These evaluations concluded that the DPIA Data

Wheel provides an effective DPIA process that any charity or organisation can use to conduct repeatable, consistent and compliant DPIAs going forward. However, the findings also suggested that more emphasis will need to be placed on differentiating between perspectives when assessing privacy risks, this will be address in the next Chapter.

The Charity who took part in this case study achieved a series of outcomes that will help them meet their obligations in respect of GDPR. These include: the MDR (Section 7.3.2) that contains a list of all the data assets (*attributes*) the Charity processes, with each attribute categorised according to sensitivity and collection justification etc. recorded as required under GDPR Article 12 (see Section 7.3.2); a new set of privacy policies that meet GDPR requirements (see Section 7.3.2); and a repeatable DPIA process (Section 7.3.4). For the wider community, this GDPR implementation presents an exemplar model for how GDPR can be implemented within a charity. However, this model can equally be applied to any small to medium enterprise (SME), or indeed organisation.

Thus, in creating the DPIA Data Wheel, it can be shown how the CLIFOD questionnaire created in Chapter 6, can be adapted to facilitate conducting a data protection impact assessment (DPIA). Further, the DPIA Data Wheel can be used and applied across any industry sector or domain, making the framework domain neutral. Thus, the DPIA Data Wheel first, provides an exemplar model for how organisations can conduct legally compliant, consistent and repeatable DPIAs going forward. Second, the DPIA Data Wheel demonstrates how contemporary legislation (GDPR) can be incorporated into the privacy assessment framework being developed as part of this thesis to practically support practitioners in privacy decision making, thereby answering RQ2, the second of the main research questions (see Section 7.1).

7.4.1 Next Steps

Based on the meta-model of CI (see Chapter 5), and the two frameworks created thus far, CLIFOD (Chapter 6) and the DPIA Data Wheel from this Chapter, the next step is to review this work to determine how these elements can be amalgamated to create one overarching framework. This will be explored in the next chapter.

Chapter 8

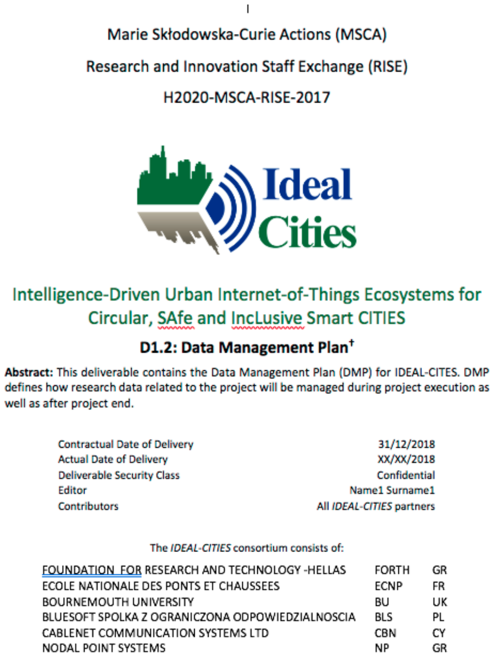
PACT Case Study

8.1 Introduction

This Chapter seeks to take the meta-model and frameworks based on CI created in Chapters 5, 6 and 7 and amalgamate these into one final framework, that will encompass all the elements discussed to create an all-encompassing privacy risk decision framework that practitioners can use to support all data privacy decisions an organisation needs to consider.

To this end an opportunity arose to take part in a Horizon 2020 European Project conducting research into how data from smart cities can be reused as open data as part of the circular economy to unlock value opportunities for citizens and businesses. *Ideal-Cities* (Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular, SAfe and IncLusive Smart CITIES) is a consortium of six organisations spanning 4 countries: France, Greece, Poland and the UK. The consortium consists of three universities: the Foundation for Research and Technology - Hellas (FORTH); Ecole des Ponts ParisTech (ENPC); and Bournemouth University (BU); and three industry partners: Bluesoft (BLS); Cablenet (CBN) and Nodalpoint (NP) (see Figure 8.1). Together these project partners (supported by a UK national charity; two local municipalities in Greece; a commune in France; and a UK Local Authority) seek to embed circular economy concepts to smart city data. What the project aims to achieve is to create an open modular platform that will take and utilise data from sensors and internet of things (IoT) technologies to, for example, facilitate mobility for disabled citizens, analysing this data to enhance their ability to move around the city safely. Further, the project also aims to offer citizens the opportunity to (re)use by making the data available in open format (thereby, contributing to a circular economy with the data) (CORDIS 2017).

The European Union (EU) is keen to reduce waste and promote reuse of resources throughout the European Economic Area (EEA) and have, as part of this drive, created an initiative to promote a *”circular economy”* (EEA 2016). This concept also relates to



Marie Skłodowska-Curie Actions (MSCA)

Research and Innovation Staff Exchange (RISE)

H2020-MSCA-RISE-2017

Ideal Cities

Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular, SAfe and IncLusive Smart CITIES

D1.2: Data Management Plan†

Abstract: This deliverable contains the Data Management Plan (DMP) for IDEAL-CITES. DMP defines how research data related to the project will be managed during project execution as well as after project end.

Contractual Date of Delivery	31/12/2018
Actual Date of Delivery	XX/XX/2018
Deliverable Security Class	Confidential
Editor	Name1 Surname1
Contributors	All *IDEAL-CITIES* partners

The *IDEAL-CITIES* consortium consists of:

FOUNDATION FOR RESEARCH AND TECHNOLOGY -HELLAS	FORTH	GR
ECOLE NATIONALE DES PONTS ET CHAUSSEES	ECNP	FR
BOURNEMOUTH UNIVERSITY	BU	UK
BLUESOFT SPOLKA Z OGRANICZONA ODPOWIEDZIALNOSCIA	BLS	PL
CABLENET COMMUNICATION SYSTEMS LTD	CBN	CY
NODAL POINT SYSTEMS	NP	GR

Figure 8.1: Ideal Cities DMP Project Partners excerpt

data and product design and how these principles can be adapted to data analytics and to promote value opportunities for citizens to become prosumers (i.e. co-producers and co-creators of the data generated from their devices into smart city use) (EEA 2017). BU, as one of the project partners, are required to focus on security and privacy. To this end this researcher was asked to contribute the privacy framework for the project.

In order to achieve this, the existing two frameworks, CLIFOD (Chapter 6) and the DPIA Data Wheel (Chapter 7) would need to be revisited to determine how these can be merged to create a comprehensive privacy risk decision framework that will support organisations in any decision making they need to make to support their business that concerns data privacy. An amalgamation of the two frameworks would, if done well, meet the requirements of the Ideal-Cities project, confirm proposition P3 (see Chapter 1, section 1.2.1) and answer RQ3 (see Chapter 1, Section 1.2).

The resulting framework, PrivACy Throughout (PACT) is presented in this Chapter. Starting with the Protocol for this case study in Section 8.2, the Chapter goes on to explain the structure of PACT by discussing each stage of creating the framework. This starts with an explanation for how CLIFOD and the DPIA Data Wheel were amalgamated in Section 8.4, followed by and explanation of how the privacy lifecycle plan (PLAN), a data governance management tool, was created in Section 8.5 to support the privacy framework. Section 8.6 explains how PACT was evaluated and Section 8.8 concludes the chapter with a summary of findings.

8.2 Protocol - Ideal Cities Case Study: Facilitating PrivACy Throughout (PACT)

8.2.1 Problem Definition

So far, two case studies have been conducted, resulting in two privacy decision making frameworks being devised based the initial meta-model of contextual integrity (CI) created in Chapter 5, CLIFOD (Chapter 6) and the DPIA Data Wheel (Chapter 7). This study will be a third case study, devised to compliment existing work and build on the concepts devised as part of these two case studies.

The aim of this study will be to review and revise the existing two frameworks and look at how these can be used and/or amalgamated to better support organisations in making informed decisions about privacy risks. As part of this a review of existing tool support will be conducted to inform an additional element of the final privacy risk framework, a privacy forward planning tool that will allow organisations plan how they will manage the data going forward. This tool will be devised to complement the existing frameworks, the DPIA Data Wheel and CLIFOD to form a the final practical framework, PrivACy Throughout (PACT), an overarching privacy decision support framework that incorporates privacy decision making into the heart of organisational decision making.

PACT will be a bridging tool, a privacy data lifecycle framework, that will allow organisations to plan how they will safeguard and manage the data going forward and review and update the privacy risk decisions made as part of completing CLIFOD and/or the DPIA Data Wheel elements. Moreover, completing the privacy data lifecycle using this tool will help demonstrate compliance with GPDR and aid decision makers in ensuring continued safeguarding of any personal data.

8.2.2 Research Questions

This chapter will answer research question 3 (RQ3) and proposition 3 (P3). This work was conducted as a case study (see Section 1.2). To recap these are:

RQ3 *How can the privacy-specific assessment framework be adapted into a tool that can support any privacy decision making the organisation need to consider?*

P3 *Can a single tool or prototype be created based on the PAF that can facilitate comprehensive privacy-specific decision making support?* (see Section 1.2).

8.2.3 Methodology

In order to answer the research questions the case study will be carried out in three stages as follows:

Stage 1 - PrivACy Throughout (PACT) - Amalgamate CLIFOD/DPIA Data Wheel results into one framework to facilitate the creation of one framework tool that can accommodate privacy decision making for both existing (CLIFOD) and new (DPIA Data Wheel) processes, systems, projects or other actions involving the handling of data. Revise both models to align with each other and facilitate smooth integration into the data lifecycle planning phase;

Stage 2 - PLAN : Creating an additional phase within PACT, a Privacy Lifecycle PlAnnINg (PLAN) tool for data governance. This PLAN will use the information gathered and decisions made in CLIFOD and/or the DPIA Data Wheel to inform and help forward plan how the data will be managed, safeguarded and stored throughout the rest of its lifecycle with the organisation. This will involve establishing how the stages of the data lifecycle can be factored in as part of any forward planning for the data. This will be based on the adaptation of the (Altman et al. 2015) data lifecycle model devised in earlier work, see Chapter 2, Figure 8.6). This will also look at how decisions can be aligned to GDPR principles and privacy goals;

Stage 3 - evaluating PACT : an evaluation will be conducted working in collaboration with one of the Ideal Cities project partner organisations to trial the PACT framework. This will involve asking the participating organisation to complete PACT for:

- An existing system, process or project using CLIFOD; and
- A new or proposed alteration to an existing system, process or project, using the DPIA Data Wheel;
- Complete the PLAN for both the assessments above.

8.2.4 Ethics Approval

Ethics approval for this case study will be sought from the University Ethics Committee, a copy of the application can be found in **??**, Section E.0.1.

8.3 Ideal Cities Case Study

In accordance with the protocol (see Section 8.2.3), this case study will consist of three stages as follows:

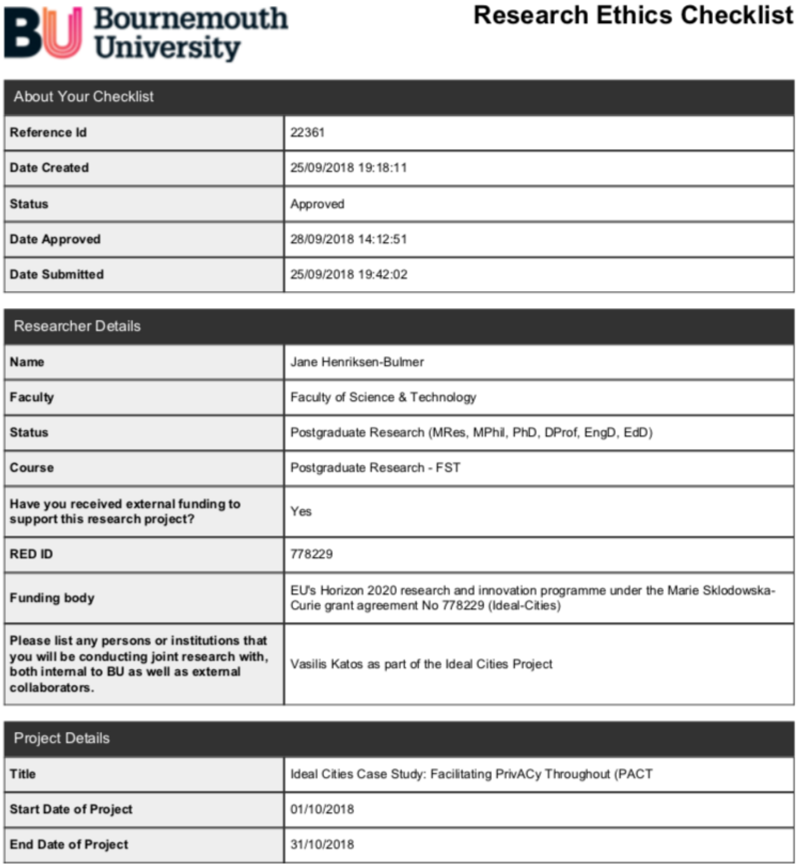
Stage 1 - PrivACy Throughout (PACT) - Amalgamate CLIFOD/DPIA Data Wheel results into one framework (Section 8.4);

Stage 2 - PLAN : Creating an additional phase within PACT, a Privacy Lifecycle PlAnnINg (PLAN) tool for data governance (Section 8.5);

Stage 3 - evaluating PACT : an evaluation will be conducted working in collaboration with one of the Ideal Cities project partner organisations to trial the PACT framework (Section 8.6).

8.3.1 Ethics

Ethics approval was sought from the university's ethics committee and granted (see Figure 8.2).



Bournemouth University

Research Ethics Checklist

About Your Checklist	
Reference Id	22361
Date Created	25/09/2018 19:18:11
Status	Approved
Date Approved	28/09/2018 14:12:51
Date Submitted	25/09/2018 19:42:02

Researcher Details	
Name	Jane Henriksen-Bulmer
Faculty	Faculty of Science & Technology
Status	Postgraduate Research (MRes, MPhil, PhD, DProf, EngD, EdD)
Course	Postgraduate Research - FST
Have you received external funding to support this research project?	Yes
RED ID	778229
Funding body	EU's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 778229 (Ideal-Cities)
Please list any persons or institutions that you will be conducting joint research with, both internal to BU as well as external collaborators.	Vasilis Katos as part of the Ideal Cities Project

Project Details	
Title	Ideal Cities Case Study: Facilitating PrivACy Throughout (PACT
Start Date of Project	01/10/2018
End Date of Project	31/10/2018

Figure 8.2: Ethics Approval - Facilitating Privacy through PACT

8.4 Stage 1 - PrivACy Throughout (PACT)

The first stage will involve taking the two frameworks created in previous work, CLIFOD and the DPIA Data Wheel and consolidating these into one spread sheet. As part of this, each framework will be revisited and reviewed and any changes made based on the outcome of the evaluations within each of the supporting case studies.

8.4.1 PACT Format

To start, a spreadsheet will be created that will include the following worksheets:

Worksheet 1 - Instructions &Definitions This worksheet will be copied from the DPIA Data Wheel with the instructions from the CLIFOD questionnaire incorporated into the appropriate sections;

Worksheet 2 - CLIFOD This will be a copy of the CLIFOD questionnaire. This will be amended to ensure changes brought in by GDPR are incorporated. Further, CLIFOD will also be revised and reviewed in accordance with identified suggested modifications from the CLIFOD case study outcomes in Chapter 6 , Section 6.6.4 and findings from the DPIA Data Wheel;

Worksheet 3 - Is DPIA Required This will be a copy of the ”Need for a DPIA” worksheet in the DPIA Data Wheel;

Worksheet 4 - DPIA DataWheel This will be a revised copy of the ”DPIA DataWheel” worksheet in the DPIA Data Wheel;

Worksheet 5 - Data Register taken from the DPIA Data Wheel;

Worksheet 6 - Life of the Form taken from the DPIA Data Wheel;

Worksheet 7- Privacy Risks taken from the DPIA Data Wheel;

Worksheet 8 - Privacy data lifecycle this will be the new phase (see Stage 2);

Worksheet 9 - List taken from the DPIA Data Wheel and updated and revised as necessary to support the final PFT.

8.4.2 CLIFOD - revision(s)

In Chapter 6 , Section 6.6.4 a number of suggestions for modifications of the CLIFOD questionnaire were made. In addition, the introduction of GDPR may have an impact on some of the questions and therefore, the this will need to be accounted for in the revised CLIFOD phase within the PACT.

Creating PrivACy Throughout (PACT) framework involved reviewing the two frameworks created in previous chapters, CLIFOD (Chapter 6) and the DPIA Data Wheel (Chapter 7) based on the meta-model (Chapter 5), to establish how these could be translated or transferred into one comprehensive privacy risk assessment framework that practitioners can use to support any privacy decision regarding data.

The first stage of this process consisted of a review of CLIFOD and the DPIA Data Wheel to determine how these two can be amalgamated into one final framework, PriVAcy

Throughout (PACT) (see Section 8.4.3). The starting point was to take the two existing frameworks and merging these into one spreadsheet. However, while some parts of both frameworks overlap and all consider context, it is appropriate to reflect the differences in focus in PACT by separating them into risks to the individual and risks to the organisation. By splitting the risks this way will ensure that any risks identified in the 'risks to the individuals' assessment, can then be assessed for how they relate to the organisation. This way, having the two risk assessments separate will then serve a second purpose that ensures appropriate organisational and/or system safeguarding measures can be identified and implemented to protect against the risks occurring.

CLIFOD assesses the privacy risks associated with sharing, publishing or otherwise making data available to third parties. The resulting list of risks from competing CLIFOD will be the risks to the organisation if data is released. The intention is that the revised framework can be used for conducting privacy assessment on ANY dataset, not just open data, to establish whether or not the dataset under consideration can be shared with third-parties or, if not, identify what steps need to be taken to make it shareable.

With the DPIA Data Wheel, while this assessment also seeks to protect sensitive and/or personal data from being shared, the perspective from which this assessment is conducted is different. The DPIA Data Wheel aims to assess what the consequences of a risk occurring would be for the *individual* whose data is compromised, rather than what the risks are to the organisation who is processing the data. As a result, the focus of the two privacy frameworks differ yet, some parts seek to capture the same information. Therefore, rather than having two separate assessments asking the same questions twice, it would save time and effort if those parts that relate to both can be answered only once which can be achieved by amalgamating the two into a single framework.

Moreover, the results of the DPIA Data Wheel evaluation showed that, despite instruction, most practitioners only considered risks from an organisational perspective (see Sections 7.3.4 and 7.4). Therefore, while amalgamating parts of the framework to avoid duplication of questions is a good idea, for the risk assessment, a distinct separation is necessary. This will help reduce confusion and highlight the differences in perspective and ensure practitioners approach each assessment from the correct perspective, the format of PACT has been changed.

Therefore, in the PACT framework, to address this, these two perspectives have been split into two separate risks assessments (Steps 4 and 5). Further, in order to really re-iterate the difference between risks to the individual and risks to the organisation, a lot more information has been provided in the risk sections. These sections therefore now include much more detailed guidance on what a risk to the individual or organisation is and how to score the risk. This includes examples what might constitute that level of severity, impact or likelihood for each score level. For this guidance, the scoring guidelines and levels used have been adapted from the PIA risk scoring methodology of the French

Data Protection Office (CNIL) (CNIL 2012) (see Appendix F, Sections F.5 and F.6).

The PACT framework will be created as a paper prototype spreadsheet (see Chapter 3, Section 3.4.11) with each step being contained on a worksheet (*'tab'*) within that spreadsheet. The explanation for each of these steps is provided in the Sections 8.4.3 to 8.4.5 and Appendix F.

8.4.3 PACT Step 1 - Overview

By analysing the similarities between CLIFOD and the DPIA Data Wheel so that these could be amalgamated into one questionnaire that will provide an overview of the data, system, process or project being assessed. To this end, a UML Activity diagram was created to depict which parts of CLIFOD and the DPIA Data Wheel align with the different aspects of contextual integrity, in particular, the three key elements: explanation, evaluation and prescription, which in CLIFOD and the DPIA Data Wheel were translated into: explanation/data, risk assessment/wheel and decision/consultation respectively.

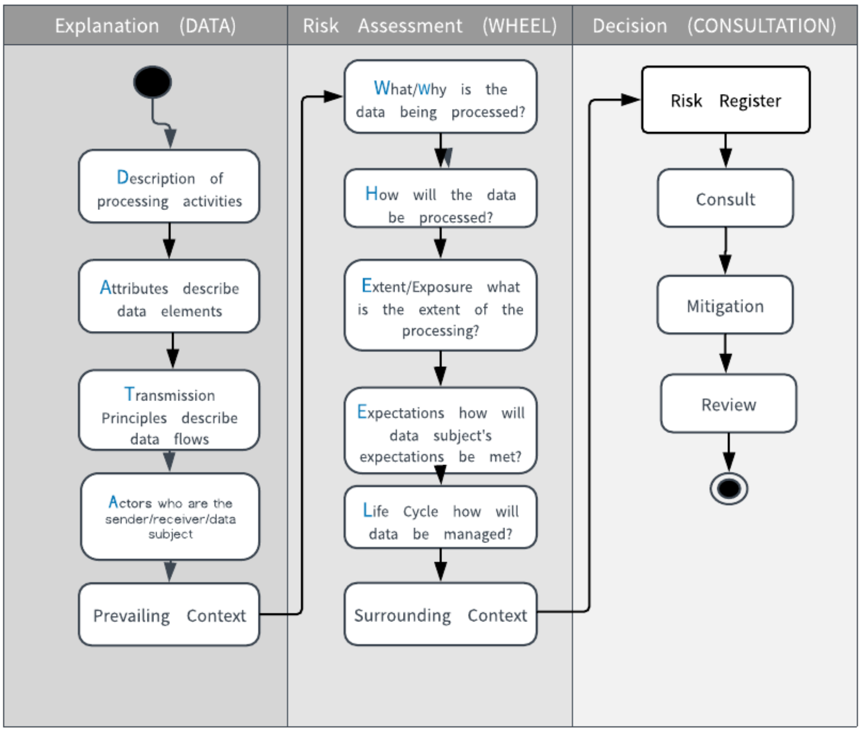

Explanation (DATA)
Risk Assessment (WHEEL)
Decision (CONSULTATION)
Description of processing activities
Attributes describe data elements
Transmission Principles describe data flows
Actors who are the sender/receiver/data subject
Prevailing Context
What/Why is the data being processed?
How will the data be processed?
Extent/Exposure what is the extent of the processing?
Expectations how will data subject's expectations be met?
Life Cycle how will data be managed?
Surrounding Context
Risk Register
Consult
Mitigation
Review

Figure 8.3: UML Activity Diagram of the Data Wheel

Figure 8.3 shows which aspects of the Data Wheel fall into each of the three phases within CLIFOD. From this activity diagram it can be seen that the explanation phase in CLIFOD becomes the 'Data' part of the DPIA Data Wheel, while the risk assessment became the 'Wheel' and the decision became the consultation phase.

Looking at the first phase in both CLIFOD and the DPIA Data Wheel (explanation/data), it is clear that, in both frameworks, the explanation phase seeks to collect the same information: details about the data; the actors; the transmission principles; and the prevailing context. However, the DPIA Data Wheel Data element, has fewer questions while ensuring the information captures aligns with GDPR. Thus, the comparison can be found in Table 8.1.

No.	**CLIFOD** (Chapter 6)	**DPIA Data Wheel** (Chapter 7)
1.	**Describe Dataset** In the original CLIFOD the first section consisted of a question which ask practitioners to *Describe* what the dataset is about. This question was not numbered as it was devised to set the scene for the rest of the questionnaire	**Description** In the DPIA Data Wheel, the corresponding 'set the scene' questions in the DPIA Data Wheel (Data part) relate to the D (Description), but with some additional questions to reflect GDPR and make sure there is a legitimate legal basis for processing the data processing
2.	**Data** The next 13 questions in the first section of CLIFOD concerned the attributes (Q1-12, the first question is not numbered)	**Attributes** The information relating to the attributes was captured in the A of Data within the DPIA Data Wheel, asking practitioners one question about what attributes will/has be/been collected and to complete the data register to capture more details
3.	**Information flows** In CLIFOD the questions around the transmission principles are in two sections. The first section consists of 4 questions (Q13-16) ask about the existing information flows. It was designed to capture the context within which the data is currently transmitted and where the data originates from (which department/unit within the organisation 'owns' the data)	
3	**Transmission Principles** The second set of questions that relate to transmission principles in CLIFOD concern how the transmission principles will change (Q48-52)	**Transmission Principles** These two sections in CLIFOD correspond with the 'T' of Data within the DPIA Data Wheel. However, the questions here are more focused on how the data is processed and shared both internally and externally
4.	**Actors** This is the largest section of the explanation phase within CLIFOD (Q17-47), asking details about the people who will be transmitting the data (*sending and receiving*) and the data subject and how these relate to each other	**Actors** In the DPIA Data Wheel, this is the second A of data, it consists of 18 questions about who the actors are and the relationships between processing actors and the data subject
5.	**Prevailing Context** This section is quite comprehensive in CLIFOD and includes some background explanation about the concepts	**Prevailing Context**The prevailing context section in the DPIA Data Wheel was copied more or less directly from CLIFOD.

Table 8.1: Comparison CLIFOD and DPIA Data Wheel

Therefore, in PACT, the explanation phase needs to become a combination of these two frameworks and renamed 'Step 1 - Overview'. The name change is necessary because in the PACT framework this step is intended to provide an overview of the scenario or processing activity being assessed, irrespective of whether personal data is collected or not. This way, the overview will apply to any privacy risk decision making that an organisation might need to consider. The intention is for PACT to be an all encompassing framework for assessing privacy risks of any new, existing or proposed processing of data. This therefore, needs to encompass any strategic or operational decision making around privacy that an organisation needs to make which could involve a number of different scenarios (e.g. a new or exiting project, system design or implementation or proposed strategic or operational change in processes). The layout and for the overview was designed based on the overarching format of 'Data' (from the DPIA Data Wheel), and the questions are based on the two existing frameworks being amalgamated as follows:

1. The questions from the DPIA will be utilised as these are more comprehensive and captures the legal obligations from GDPR. However, there will be some slight amendments with the questions being rephrased slightly (all the questions are listed in order in Appendix F, Section F.2);

2. The evaluation of the CLIFOD framework revealed that this section of questions in the questionnaire did not flow very well, the format was confusing some of the questions were surplus to requirements for many of the attributes and not gathering enough detail for others. This was addressed in the A of Data within the DPIA Data Wheel by asking practitioners just one questions and then asking them to complete the data register to capture the detail and ensure the information required to comply with GDPR is captured, this format will be retained but the data register will be amended to facilitate use for any privacy risk assessment rather than just the DPIA, the amended list of categories on the Data Register will be described in Appendix F, Section F.2.1;

3. For the transmission principles, the feedback from the evaluation on the DPIA showed that the questions asked do not capture enough detail about the prevailing context. For CLIFOD the questions were geared towards data publication and how the changes in data flow might impact on privacy to cause a privacy violation. Moreover, both sets of questions require some explanation for the novice assessor or practitioner to capture the detail required. Therefore, this section has been revised to first, add an additional explanation column to the PACT questionnaire. Second, taking a combination of the questions from both frameworks, rephrasing and adding some more questions to make this section clearer and easier for practitioners to complete (see Q7-18 in Appendix F, Section F.2);

4. The actors section will be a combination of the two existing frameworks with some additional questions added for more detail and clarification;

5. The life of the form from the DPIA Data Wheel will be reused in PACT to inform the privacy risk assessment, the description for this worksheet can be found in Appendix F, Section F.2.2);

6. The final section captures the prevailing context. This section has been revised from the previous version used in the two frameworks slightly with some additional questions and slight rephrasing of the existing questions to make them apply to assessments for either new or existing practices.

The amalgamated questionnaire consists of 50 questions as follows:

Q1 Describe the data/dataset; What is the system, process, project or dataset about (subject matter/context)?

Q2 What is the purpose of the data collection/processing?

Q3 What is the lawful basis for collecting/processing the data (drop-down list). see Chapter 2, Section 2.4.4 for a list of these)

Guidance Q3 GDPR places a requirement on organisations to specify a lawful basis for processing of personal data.

Q4 If special category data processed, please also provide secondary lawful basis for processing the data (drop-down list). see Chapter 2, Section 2.4.4 for a list of these) Where special category data (e.g. data relating to health) is processed, a secondary lawful basis for processing may also be selected e.g. the primary lawful basis may be contractual with consent as a secondary lawful basis for certain elements.

Guidance Q1-4 This section sets the scene for the assessment and provides an overview of what data/system/process or project being assessed. *Processing* refers to ”any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction”.

Q5 Attributes - at a high level, please describe what data will/has be/been collected?

Q6 Please go to the data register worksheet and complete the data register with more detailed information about the attributes that will be collected/processed. If no personal or sensitive data will be collected/processed or stored, you may decide to complete section A only (however, completing section B is recommended for all data).

Guidance Q5-6 In order to understand the information flows it is necessary to first understand the data. What is the data about and what are the individual parts that make up the datasets. *Attributes* are the individual data elements within a data set. (see Section F.2.1 for column headers in the Data Register)

Q7 If the data/system/process/project being assessed is a a new practice please describe the data flows and how it is proposed the data will flow between actors? (i.e. any stakeholders using or processing the data)

Q8 If the data/system/process/project being assessed is a change to existing practice please describe the proposed changes in data flows and what changes in flow this will bring about?

Q9 If the data/system/process/project being assessed is a new practice please describe the data flows and how it is proposed the data will flow between actors? (i.e. any stakeholders using or processing the data)

Guidance Q7-9 This section seeks to find out about data flows, i.e. how the data is processed. Please answer the question that relates to the scenario being assessed i.e. Q7 for existing; Q8 for a change to existing process/practice; or Q9 for a new process/practice.

Q10 What is the overriding context within which this processing/collection of data takes place? e.g. public service, national telecoms, health care etc.

Q11 what is the sector context nested within the overriding context where this data collection practice takes place? e.g. library, telecommunications, teaching hospital.

Q12 What is the local (or departmental) context within which the data is processed/collected? e.g. provision of bin collection services, order processing, marketing, patient records system etc.

Guidance Q10-12 This relates to familiar level of context and the embedded context within that overriding context.

Q13 How is/will the data be processed internally within the organisation?

Q14 Please describe the information flows with internal stakeholders

Q15 Will the data be transmitted to external stakeholders?

Q16 If yes, please describe the information flows with external stakeholders.

Q17 Please go data journey worksheet and complete this to gather more detailed information about the data lifecycle of the collection/project/system being assessed (see Section F.2.2 for questions in the Data Journey).

Guidance Q13-17 These questions relate to how the existing information flows, i.e. how the data is currently processed.

Q18 Who is the designated data controller for the data? (normally this will be the organisation).

Guidance Q18 GDPR requires that organisations keep a record of who is DC.

Q19 Responsible Person - name and contact details of the designated person acting on behalf of the Data Controller(s).

Guidance Q19 This will be the contact person who acts on behalf of the DC.

Q20 Who is the Data Owner? (if different from the Data Controller).

Guidance Q20 In some instances, your organisation may may process data belonging to another entity.

Q21 Name & contact details of Data Protection Officer (DPO) (if applicable).

Q22 Who is the Data Subject(s)?

Q23 What is the relationship between the Data Subject and the Data Controller/Processor/DPO? (e.g. friend, co-worker, professionally, citizen, employment) Please note all relationships that apply.

Q24 Are any internal stakeholders also Data Subject(s)? (e.g. staff).

Q25 If there are internal data subjects, please describe the relationship between them and the data owner/controller and/or DPO?

Q26 How does the Responsible person(s)/Data Owner and/or DPO interact with the internal data subject(s)? (e.g. friend, co-worker, professionally, citizen, employment) Please note all relationships that apply.

Guidance Q18-26 Please note who the actors are. These are the stakeholders. These include the data controller (DC), Data Owner (DO), Data Protection Officer (DPO), Data Processor (DP), Joint Data Controller (JC) and the data subject(s) (i.e. the individuals whose data is being processed). Further details and descriptions are available in the instructions worksheet (under actors).

Q27 Please list any Data Processors who will be processing the data on behalf of the Data Controller (e.g. Cloud based software suppliers or other contracted suppliers who may hold or process the data for or on behalf of the Controller).

Q28 What is the relationship between the Data Controller and the Data Processor? (e.g. professional/friends/employer/employee etc.)

Q29 Is a contract in place between the Data Controller and the Data Processor?

Guidance Q27-29 Data Processors (DP) please note details of any Data Processor(s) or Joint-Data Controller(s) who will process/use/access the data for or on behalf of the Data Controller and what their relationship is with the organisation. This is relevant where you use the services of an outside organisation to process data on your behalf.

Q30 Please list any Joint Data Controller(s) (e.g. external partner organisation) who will be processing the data on behalf of or in place of the Data Controller.

Q31 Is a contract in place between the Joint Data Controller and the Data Processor?

Q32 What is the relationship between the Joint Data Controller and the Data Controller? (e.g. professional/friends/employer/employee etc.)

Guidance Q30-32 Joint Data Controller(s) (JC) This is relevant where you process data jointly with another entity.

Q33 External Responsible person - name and contact details of the designated External Data Controller/Processor.

Q34 Name & contact details of third-party Data Protection Officer (DPO) if applicable.

Q35 How does the Data Processor and/or Joint Data Controller interact with the data subject(s)? (e.g. friend, co-worker, professionally, citizen, employment) Please note all relationships that apply.

Guidance Q35 DP/JC Relationship to Data Subject.

Q36 Other Actors not mentioned (please explain role(s) and relationship).

Guidance Q36 Please note any other internal or external stakeholders who will use/access the data and what their role is.

Guidance Q33-35 Please note the contact person(s) within the external stakeholders who is responsible for the data use/access and what their role is within the organisation.

Q37 What was/is the social context of the data collection e.g. hospital would be health context, school would be educational context etc.

Q38 Is/was data collected directly from data subject?

Q39 If not, please note from whom the data was obtained and whether the data subject has been informed?

Guidance Q37-39 What are the existing informational norms with regards to the data, i.e. what is the prevailing social context.

Q40 Has/will permission/consent been/be sought for processing of the data from data subject(s)?

Q41 If consent has not been obtained are there any overriding considerations as to why processing should be allowed despite lack of consent/limited consent?

Q42 Was/is the data collected with a view to process beyond its original purpose? (if yes, please answer questions below)

Q43 Does data originate from third party source (if so, please state origin of the data)

Q44 Was consent sought from the data subject for secondary processing purposes?

Q45 Were there any limitations on secondary processing purpose(s) to which consent was given?

Q46 If no consent obtained for secondary processing, are there any overriding considerations as to why secondary processing should be allowed despite lack of consent/limited consent?

Q47 Has/will the data subject been/be informed of their rights to withdraw consent?

Guidance Q40-47 *Prevailing context* refers to the existing context, for example, if an existing dataset is being assessed, what is the current context within which data is processed? For this both the local context (e.g. within the organisation, say a telecommunications company) and the surrounding context (e.g. the national telecommunications network). Locate applicable entrenched informational norms and identify significant points of departure. Consider legal and ethical obligations e.g. consent, is it required/will it be obtained?

Guidance Q37-47 This section seeks to establish the existing context within which data is collected, processed, stored or shared.

Q48 Based on the contexts identified above, are there any potential impacts from how the data flows that could result in a violation of the privacy of the *data subject*?

Guidance Q48 This question seeks to establish whether *prima facie* the existing processes and data flow could have potential to violate the privacy of the data subject.

Q49 If the data/system/process/project being assessed is a change to existing practice (e.g. contracted third party processing on your behalf) please describe the proposed changes in data flows this change will bring about.

Q50 Based on the contexts identified above, *prima facie* is there any potential impacts from the proposed change in data flows that could result in a violation of *privacy of the data subject* or *compromise the confidentiality of the organisation*?

Guidance Q49-50 These question seek to establish whether, in light of any third-party data processing (DP contractors and/or JC's create a change in existing processes and data flows that could have potential to violate privacy.

8.4.4 PACT Step 2 - Need for a DPIA

Step 2 of the framework involves deciding whether a DPIA is needed (see Appendix F, Section F.3 for details of the questions on this worksheet (*'tab'*). Initially, this was placed as Step 3 in the framework but after some deliberation, it was moved to Step 2 to allow for the outcome of this assessment to feed into Step 3, the wider context. This way, when practitioners get to Step 3, they already know whether or not they need to assess risks to the individual and this will influence the answers provided there, these questions are are a the same as those used in ”the need for a DPIA” assessment in the DPIA Data Wheel (see Section 7.3.4).

8.4.5 PACT Step 3 - Wider Context

Step 3 forms the first part of the Risk Assessment (evaluation in terms of CI). This step consists of the context questions from CLIFOD and the DPIA Data Wheel with some additional guidance and clarification questions. The reason for moving the context to the beginning of the risk assessment is twofold. First, where the outcome of Step 2 is that a DPIA is not required, practitioners will not need to complete Step 4, the 'Risks to the Individuals' section. However, irrespective of whether practitioners need to do a DPIA or not, they will still need to consider the surrounding (wider) context as this can impact on the data flows. For example, the staff who process the data might have access to personal information as part of their role, however, if they were to share this information on social media, they would have breached a number of contextual norms and values including; breaching the data subject's trust and taken a prerogative in divulging the information and, of course, infringed on the legal obligation to preserve confidentiality.

It was therefore decided that the easiest way to accommodate this was to separate the surrounding context questions from the two original frameworks out and place these as the first step of the risk assessment. This ensures that first, the wider context is considered and second, by having this as the first step of the risk assessment, the answers provided here can be used to inform the privacy risk assessment, thereby incorporating context at an earlier stage in the evaluation. Second, moving the context questions to the beginning of the risk assessment also serves to raise the prominence of the need to consider risks in context with surrounding factors, which is the main aim of CI. The questions this section asks are:

Q51 How could data collection, processing, transmission or publication be perceived to infringe on any political values? (please also explain whose values and how they might be infringed upon)

Norm/Value Q51 Political

Guidance Q51 Values are the underlying standards and specific guidelines that we apply to the norms. They are standards or principles of behaviour i.e. our judgement of what is important. Our values are what tells us what is good or bad, what is acceptable and what is not. Values include democracy, freedom of speech and autonomy. In terns of data, our values, societal and individual, will influence how we approach and handle the data which may, for example, be in the best interest for profit but not for the individual whose data is being processed.

Q52 How might data processing be perceived to impose any form of power imbalance? (please explain how such a power balance might arise, who might be affected and how these might be imposed)

Norm/Value Q52 Power imbalance

Q53 How could data processing infringe on any legal obligations? (please clarify what legal compliance might be breached, how and who would be affected)

Norm/Value Q53 Legal

Guidance Q52-53 Values: What values could external or internal stakeholders potentially abuse or be perceived to abuse? Are there any factors, influences or aspects that could be perceived to support/discourage data collection, processing, transmission or publication e.g. what might be the potential effects on implications for justice, power structures, threat to democracy etc.

Q54 How might data processing be perceived to infringe upon any social values? (please explain what those values are, whose values they are and how these might be infringed upon)

Norm/Value Q54 Social

Q55 How might data processing be perceived to infringe any moral/ ethical values? (please explain what those moral values are, whose values and how these might be infringed upon)

Norm/Value Q55 Moral/Ethical

Guidance Q54-55 Are there any ethical factors that may support/discourage data processing e.g. would data collection, processing, transmission or publication threaten fairness, equality or social hierarchy

Norm/Value Q56 Belief system(s)

Q57 How might data processing pose a threat to the autonomy or freedom of the data subject(s)? (please explain how this might pose a threat, who could be affected and how)

Norm/Value Q57 Autonomy & Freedom

Q58 In what ways could data processing result in informational harm on the data subject(s)? (please clarify how such harm could occur, who would be affected and how)

Norm/Value Q58 Harm

Q59 How might data processing give rise to any form of discrimination (please explain what form of discrimination, how this might occur and who would might be adversely affected)

Norm/Value Q59 Discrimination

Q60 In what ways might data processing grant or afford any privileges or prerogatives to the referrer, the data- processor(s), controller(s) or originator(s) in a manner that benefits or is perceived to benefit those parties as a result of data processing?

Norm/Value Q60 Privileges or Prerogatives

Q61 In what ways might data processing result in breach of trust (if yes, please clarify how such harm could occur, who would be affected and how)How could data processing impose security risks on any stakeholders (please clarify what the security risk is, how and who would be affected)

Norm/Value Q61 Security

Q62 In what ways might data processing result in breach of trust (if yes, please clarify how such harm could occur, who would be affected and how)

Norm/Value Q62 Trust

Q63 In what ways could the data processing result in a breach of confidentiality? (please clarify how such a breach might occur, who would be affected and how)

Norm/Value Q63 Confidentiality

Q64 Is there a reasonable expectation on the part of the data subject(s) and/or data controller(s) of data being kept confidential and not collected, processed, transmitted or published?

Norm/Value Q64 Confidentiality

Guidance Q56-64 Norms are customs or expected rules, expectations or standards for how we are supposed to behave in society. Norms can relate to behaviours, ideas and beliefs etc. that have become embedded within us as individuals and society, i.e. influences and standards of expected behaviours or customs. Norms affect how we react, behave and perceive the world around us. Examples of norms include political or religious beliefs, an expectation of non-discrimination, equality (e.g. equal opportunity etc.) Norms: What norms could external or internal stakeholders potentially infringe upon or be perceived to infringe upon?

Q65 What is/are the desired effects of the processing for the data subject?

Q66 What are the benefits and positive values for the data subject that processing the data will bring to enhance the outcome for the data subject e.g. to facilitating treatment, supply goods, improve mobility, increase transparency etc.

Norm/Value Q65-66 Data Subject

Q67 What is the desired effect of the processing for the organisation?

Q68 What are the benefits and positive values that data processing will bring/enhance for the organisation. These may include commercial gain, improved transparency, meeting legal obligation etc.

Norm/Value Q67-68 Organisation

Guidance Q65-68 Positive Impact Assessment: note any positive values that could be derived from the data collection, processing, transmission or publication?

Q69 If consent has not been obtained from the data subject, or they have withdrawn their consent, are there any overriding legal, moral or ethical reasons why data processing should be allowed despite this? (e.g. prevention of terrorism, safeguarding of the data subject etc.)

Q70 Are there any overriding legal, moral or ethical reasons why data processing should be allowed even if there is a risk of re-identification?

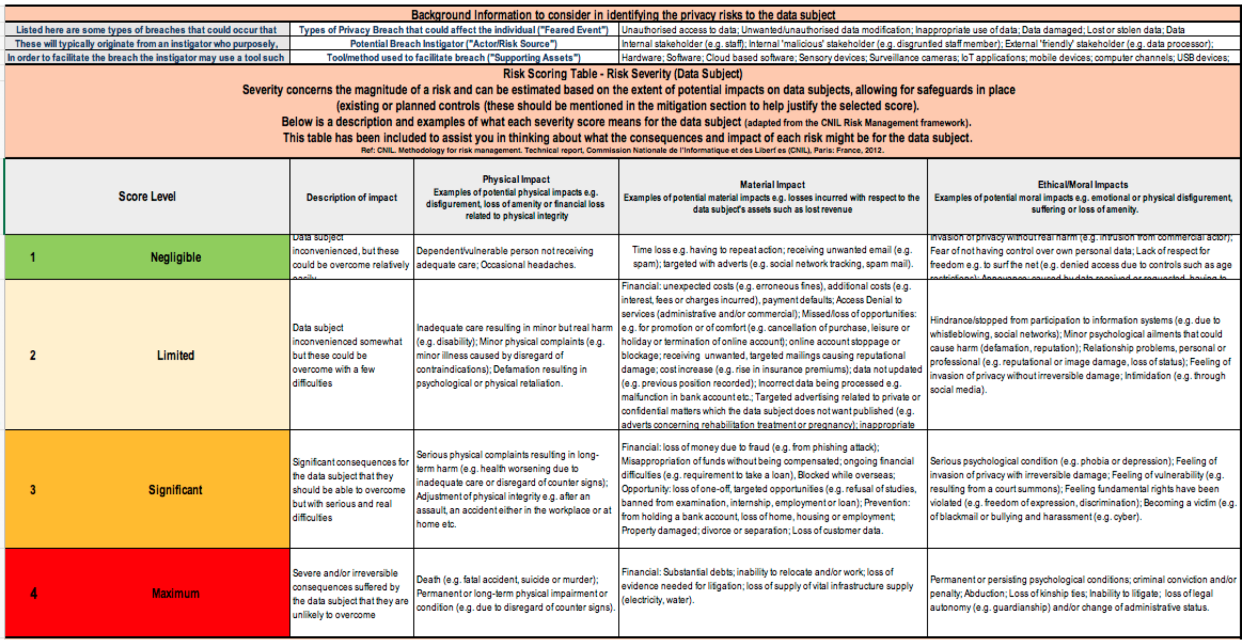
Norm/Value Q69-70 Organisation

Guidance Q69-70 Overriding reason for processing despite risks or lack of consent.

8.4.6 PACT Step 4-Risks to the Individual

Step 4 is the Data Protection Impact Assessment (DPIA), the second part of the risk assessment. This is named 'Risks to the Individual' to first, highlight that the risks here should be considered from the perspective of the individual rather than the organisation. Second, practitioners may decide to conduct this assessment irrespective of whether any *"high risk"* processing will be carried out. GDPR only stipulates that a DPIA is required where *"high risk"* processing is conducted and one of the following apply: implementing new technologies; processing involves extensive profiling or automated processing, genetic/biometric data or large scale processing of sensitive data; extensive systematic monitoring; use of automated decision-making or profiling; involves data linking, evaluation/scoring, tracking, children's data or could cause potential harm to the data subject (a full list of these can be found in Chapter 2, Section 2.4.4).

As mentioned in Section 8.4, to help practitioners understand the difference between the perspectives of the individual and the organisation when assessing the risks, the guidelines for examples of breach types, who might instigate the breach and how to score any risks identified have been adapted from the CNIL methodology (CNIL 2012), an excerpt can be found in Figure 8.4, while the full wording, explanations and column headers etc. for this step can be found in Appendix F, Section F.5.



Background Information to consider in identifying the privacy risks to the data subject

Listed here are some types of breaches that could occur that	Types of Privacy Breach that could affect the individual ("Feared Event")	Unauthorised access to data; Unwanted/unauthorised data modification; Inappropriate use of data; Data damaged; Lost or stolen data; Data
These will typically originate from an instigator who purposely,	Potential Breach Instigator ("Actor/Risk Source")	Internal stakeholder (e.g. staff); Internal 'malicious' stakeholder (e.g. disgruntled staff member); External 'friendly' stakeholder (e.g. data processor);
In order to facilitate the breach the instigator may use a tool such	Tool/method used to facilitate breach ("Supporting Assets")	Hardware; Software; Cloud based software; Sensory devices; Surveillance cameras; IoT applications; mobile devices; computer channels; USB devices;

Risk Scoring Table - Risk Severity (Data Subject)
Severity concerns the magnitude of a risk and can be estimated based on the extent of potential impacts on data subjects, allowing for safeguards in place
(existing or planned controls (these should be mentioned in the mitigation section to help justify the selected score).
Below is a description and examples of what each severity score means for the data subject (adapted from the CNIL Risk Management framework).
This table has been included to assist you in thinking about what the consequences and impact of each risk might be for the data subject.
Ref: CNIL. Methodology for risk management. Technical report, Commission Nationale de l'Informatique et des Libert es (CNIL), Paris: France, 2012.

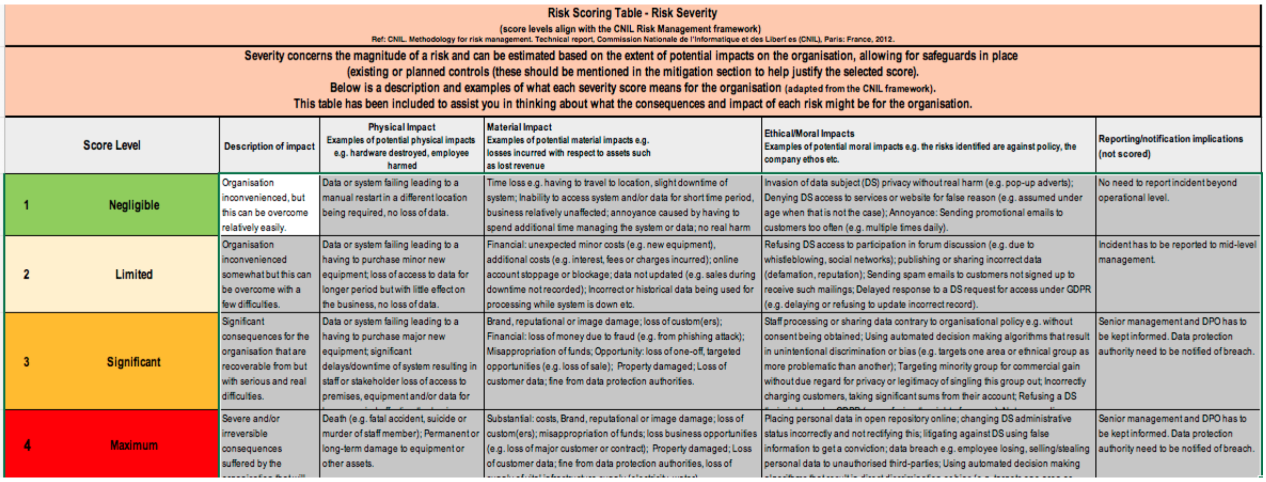
Score Level		Description of impact	Physical Impact Examples of potential physical impacts e.g. disfigurement, loss of amenity or financial loss related to physical integrity	Material Impact Examples of potential material impacts e.g. losses incurred with respect to the data subject's assets such as lost revenue	Ethical/Moral Impacts Examples of potential moral impacts e.g. emotional or physical disfigurement, suffering or loss of amenity.
1	Negligible	Data subject inconvenienced, but these could be overcome relatively easily	Dependent/vulnerable person not receiving adequate care; Occasional headaches.	Time loss e.g. having to repeat action; receiving unwanted email (e.g. spam); targeted with adverts (e.g. social network tracking, spam mail).	Invasion of privacy without real harm (e.g. intrusion from commercial actor); Fear of not having control over own personal data; Lack of respect for freedom e.g. to surf the net (e.g. denied access due to controls such as age restrictions); Annoyance caused by data received or requested, having to
2	Limited	Data subject inconvenienced somewhat but these could be overcome with a few difficulties	Inadequate care resulting in minor but real harm (e.g. disability); Minor physical complaints (e.g. minor illness caused by disregard of contraindications); Defamation resulting in psychological or physical retaliation.	Financial: unexpected costs (e.g. erroneous fines), additional costs (e.g. interest, fees or charges incurred), payment defaults; Access Denial to services (administrative and/or commercial); Missed/loss of opportunities: e.g. for promotion or of comfort (e.g. cancellation of purchase, leisure or holiday or termination of online account); online account stoppage or blockage; receiving unwanted, targeted mailings causing reputational damage; cost increase (e.g. rise in insurance premiums); data not updated (e.g. previous position recorded); Incorrect data being processed e.g. malfunction in bank account etc.; Targeted advertising related to private or confidential matters which the data subject does not want published (e.g. adverts concerning rehabilitation treatment or pregnancy); inappropriate	Hindrance/stopped from participation to information systems (e.g. due to whistleblowing, social networks); Minor psychological ailments that could cause harm (defamation, reputation); Relationship problems, personal or professional (e.g. reputational or image damage, loss of status); Feeling of invasion of privacy without irreversible damage; Intimidation (e.g. through social media).
3	Significant	Significant consequences for the data subject that they should be able to overcome but with serious and real difficulties	Serious physical complaints resulting in long-term harm (e.g. health worsening due to inadequate care or disregard of counter signs); Adjustment of physical integrity e.g. after an assault, an accident either in the workplace or at home etc.	Financial: loss of money due to fraud (e.g. from phishing attack); Misappropriation of funds without being compensated; ongoing financial difficulties (e.g. requirement to take a loan), Blocked while overseas; Opportunity: loss of one-off, targeted opportunities (e.g. refusal of studies, banned from examination, internship, employment or loan); Prevention: from holding a bank account, loss of home, housing or employment; Property damaged; divorce or separation; Loss of customer data.	Serious psychological condition (e.g. phobia or depression); Feeling of invasion of privacy with irreversible damage; Feeling of vulnerability (e.g. resulting from a court summons); Feeling fundamental rights have been violated (e.g. freedom of expression, discrimination); Becoming a victim (e.g. of blackmail or bullying and harassment (e.g. cyber).
4	Maximum	Severe and/or irreversible consequences suffered by the data subject that they are unlikely to overcome	Death (e.g. fatal accident, suicide or murder); Permanent or long-term physical impairment or condition (e.g. due to disregard of counter signs).	Financial: Substantial debts; inability to relocate and/or work; loss of evidence needed for litigation; loss of supply of vital infrastructure supply (electricity, water).	Permanent or persisting psychological conditions; criminal conviction and/or penalty; Abduction; Loss of kinship ties; Inability to litigate; loss of legal autonomy (e.g. guardianship) and/or change of administrative status.

Figure 8.4: Excerpt of guidance - Risk to the Individual

8.4.7 PACT Step 5 - Risks to the Organisation

Step 5 is the third, and final, part of the risk assessment. Here practitioners are asked to consider privacy risks from the organisational perspective. The format and layout of this worksheet (*'tab'*) within the spreadsheet is very similar to the layout for the risks to the individual outlined in Step 4 in Section 8.4.6 above, but with the emphasis on the organisation rather than the individual. Privacy risks from the perspective of the organisation are as important, if not more so, as risks to the individual. In order for the organisation to safeguard against any risks identified in Step 4, they will need to identify how these relate to the organisation, so that they can address and mitigate the risks appropriately. Further, as mentioned in the Step 4 guidance (see Appendix F, Section F.5), there may be occasion where a DPIA (risks to the individual) is not required. However, the privacy risks to the organisation will still need to be assessed.

For this step, as with the risks to the individual assessment in Step 4 (Section 8.4.6 and F.5), examples of what each severity score means for the organisation have been provided in the guidance on this worksheet (*'tab'*). These have been changed from the



Risk Scoring Table - Risk Severity
(score levels align with the CNIL Risk Management framework)
Ref: CNIL. Methodology for risk management. Technical report, Commission Nationale de l'Informatique et des Liberf es (CNIL), Paris: France, 2012.

Severity concerns the magnitude of a risk and can be estimated based on the extent of potential impacts on the organisation, allowing for safeguards in place
(existing or planned controls (these should be mentioned in the mitigation section to help justify the selected score).
Below is a description and examples of what each severity score means for the organisation (adapted from the CNIL framework).
This table has been included to assist you in thinking about what the consequences and impact of each risk might be for the organisation.

Score Level		Description of impact	Physical Impact Examples of potential physical impacts e.g. hardware destroyed, employee harmed	Material Impact Examples of potential material impacts e.g. losses incurred with respect to assets such as lost revenue	Ethical/Moral Impacts Examples of potential moral impacts e.g. the risks identified are against policy, the company ethos etc.	Reporting/notification implications (not scored)
1	Negligible	Organisation inconvenienced, but this can be overcome relatively easily.	Data or system failing leading to a manual restart in a different location being required, no loss of data.	Time loss e.g. having to travel to location, slight downtime of system; Inability to access system and/or data for short time period, business relatively unaffected; annoyance caused by having to spend additional time managing the system or data; no real harm	Invasion of data subject (DS) privacy without real harm (e.g. pop-up adverts); Denying DS access to services or website for false reason (e.g. assumed under age when that is not the case); Annoyance: Sending promotional emails to customers too often (e.g. multiple times daily).	No need to report incident beyond operational level.
2	Limited	Organisation inconvenienced somewhat but this can be overcome with a few difficulties.	Data or system failing leading to a having to purchase minor new equipment; loss of access to data for longer period but with little effect on the business, no loss of data.	Financial: unexpected minor costs (e.g. new equipment), additional costs (e.g. interest, fees or charges incurred); online account stoppage or blockage; data not updated (e.g. sales during downtime not recorded); Incorrect or historical data being used for processing while system is down etc.	Refusing DS access to participation in forum discussion (e.g. due to whistleblowing, social networks); publishing or sharing incorrect data (defamation, reputation); Sending spam emails to customers not signed up to receive such mailings; Delayed response to a DS request for access under GDPR (e.g. delaying or refusing to update incorrect record).	Incident has to be reported to mid-level management.
3	Significant	Significant consequences for the organisation that are recoverable from but with serious and real difficulties.	Data or system failing leading to a having to purchase major new equipment; significant delays/downtime of system resulting in staff or stakeholder loss of access to premises, equipment and/or data for	Brand, reputational or image damage; loss of custom(ers); Financial: loss of money due to fraud (e.g. from phishing attack); Misappropriation of funds; Opportunity: loss of one-off, targeted opportunities (e.g. loss of sale); Property damaged; Loss of customer data; fine from data protection authorities.	Staff processing or sharing data contrary to organisational policy e.g. without consent being obtained; Using automated decision making algorithms that result in unintentional discrimination or bias (e.g. targets one area or ethnical group as more problematic than another); Targeting minority group for commercial gain without due regard for privacy or legitimacy of singling this group out; Incorrectly charging customers, taking significant sums from their account; Refusing a DS	Senior management and DPO has to be kept informed. Data protection authority need to be notified of breach.
4	Maximum	Severe and/or irreversible consequences suffered by the	Death (e.g. fatal accident, suicide or murder of staff member); Permanent or long-term damage to equipment or other assets.	Substantial: costs, Brand, reputational or image damage; loss of custom(ers); misappropriation of funds; loss business opportunities (e.g. loss of major customer or contract); Property damaged; Loss of customer data; fine from data protection authorities, loss of	Placing personal data in open repository online; changing DS administrative status incorrectly and not rectifying this; litigating against DS using false information to get a conviction; data breach e.g. employee losing, selling/stealing personal data to unauthorised third-parties; Using automated decision making	Senior management and DPO has to be kept informed. Data protection authority need to be notified of breach.

Figure 8.5: Excerpt of guidance - Risk to the Organisation

DPIA risk scores (see Appendix 7.3.4, Section 7.3.4) to mirror the format used in Step 4 (see Figure 8.5), the risks to the individual assessment. Therefore, although not part of the CNIL framework, the guidance and scoring method etc. have been devised to align in style and format with the risk scores and style as found in the CNIL methodology (CNIL 2012). The layout, guidance and questions for this worksheet (*'tab'*) can be found in Appendix F, Section F.6.

8.5 Stage 2 & PLAN Step 6

This stage is new to the PACT framework. It has been devised to provide practitioners with a data governance tool for managing the data during its lifecycle with the organisation, the privacy lifecycle plan (PLAN). This will be incorporated into PACT as the final step (Step 6) so that the PLAN facilitates privacy throughout by complimenting the risk assessment with a data governance tool that facilitates planning how data privacy will be managed during its lifecycle with the organisation.

8.5.1 Privacy Data Life Cycle

In creating the PLAN, the data lifecycle model created by (Altman et al. 2015) , was used as the starting point for creating a structure and process for the PLAN (Section 8.5.1). This lifecycle model identified each of the data lifecycle stages where a public body need to implement security controls to safeguards data privacy. To recap, this model talks about 5 lifecycle stages, referring to data: collection, transformation, retention, access and/or release and post-access (Altman et al. 2015). This model was initially adapted in earlier work to describe what types of decisions need to be made by organisations during each stage of the data lifecycle (see Figure 8.6, and Chapter 2, Section 2.4.4).

8.5.2 Data Lifecycle and organisational decision making

To recap, this model talks about 5 lifecycle stages, referring to data: collection, transformation, retention, access and/or release and post-access (Altman et al. 2015). For example, early in the data lifecycle, when data is being collected, the decisions that need to be made will be focused on informing the data subject about what the purpose of the data collection is, how the data will be used and ensuring consent has been obtained. Later in the life cycle however, the decisions that need to be made may focus on whether or not to destroy or retain the data.

Using the adaptation of this model created to organisational decision making devised in Chapter 2 to describe what types of decisions need to be made by organisations during each stage of the data lifecycle (see Figure 8.6).

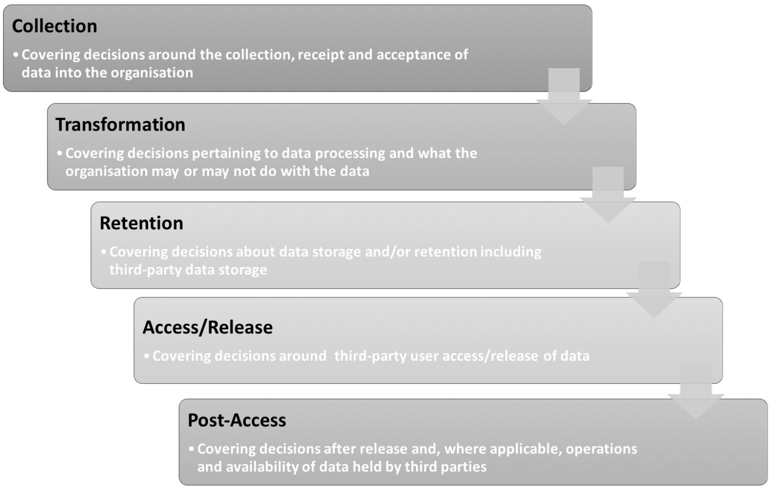

Collection
• Covering decisions around the collection, receipt and acceptance of data into the organisation
Transformation
• Covering decisions pertaining to data processing and what the organisation may or may not do with the data
Retention
• Covering decisions about data storage and/or retention including third-party data storage
Access/Release
• Covering decisions around third-party user access/release of data
Post-Access
• Covering decisions after release and, where applicable, operations and availability of data held by third parties

Figure 8.6: Adaptation of Altman et al. (2015)'s data lifecycle model for decision-making

In revisiting the adapted Data Lifecycle in Figure 8.6, it became apparent that to facilitate data management in all stages of it life with the organisation, some additional stages would need incorporating to account for data disposal and governance. Therefore, two further stages were added to reflect this resulting in a 7-stage Data Life Cycle (7-DL) for managing data throughout its life with an organisation as follows:

Collection this covers decisions regarding collection, receipt and acceptance of data;

Transformation covers decisions pertaining to data access, processing and use and restrictions placed on the data handling (e.g. legal restrictions);

Retention covers decisions about data retention and storage including time stored and third party data storage;

Access/Release covers decisions pertaining to third-party data access, processing and use and any restrictions placed (or to be placed) on such data handling (e.g. contractual, policy and/or legal);

Post-Access covers decisions regarding post-transmission and/or sharing of data and availability of data held by third-parties (adapted from (Altman et al. 2015));

Disposal covers decisions about data disposal and/or destruction including third party data disposal;

Governance & Consultation this covers consultation with stakeholders and strategic decisions that overarch all the stages of the data lifecycle e.g. decisions relating to overarching policies.

Furthermore, a model was created to depict how these stages flow, this can be found in Figure 8.7.

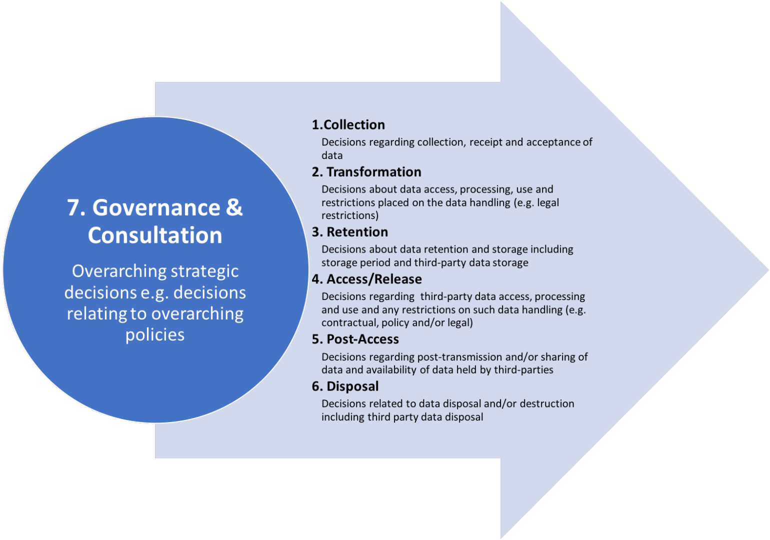

7. Governance & Consultation
Overarching strategic decisions e.g. decisions relating to overarching policies
1.Collection
Decisions regarding collection, receipt and acceptance of data
2. Transformation
Decisions about data access, processing, use and restrictions placed on the data handling (e.g. legal restrictions)
3. Retention
Decisions about data retention and storage including storage period and third-party data storage
4. Access/Release
Decisions regarding third-party data access, processing and use and any restrictions on such data handling (e.g. contractual, policy and/or legal)
5. Post-Access
Decisions regarding post-transmission and/or sharing of data and availability of data held by third-parties
6. Disposal
Decisions related to data disposal and/or destruction including third party data disposal

Figure 8.7: PLAN 7-stage Data Life Cycle, adapted from Altman et al. (2015)

Within the PLAN, the questions from the 'Wheel' of the DPIA Data Wheel have been revised and integrated, some further questions have been added. As part of this review of the DPIA Data Wheel questions, the 6 GDPR Principles were revisited to ensure these have all been covered and considered as part of the Data Life Cycle PLAN (see Chapter 2, Section 2.4.4). This highlighted that more detail would be needed to cover the security aspect of safeguarding the data.

8.5.3 Privacy Goals

Looking at the literature and existing frameworks for embedding privacy preservation into design, a common method for achieving this is through privacy goal modelling. Privacy goal modelling was discussed briefly in Chapter 2, Section 2.7.5. Similarly, for the purpose of the PLAN, it seemed an appropriate method for embedding privacy preservation into the data life cycle.

Thus, as part of creating a PLAN for how to safeguard the data during its lifecycle with the organisation, it makes sense to incorporate the most common or widely used privacy goals so that practitioners can ensure these are protected against as part of their data lifecycle management planning. To this end, a number of frameworks were identified as being relevant to privacy risk and protection. These were analysed to establish which privacy goals were most widely used or discussed, resulting in the following frameworks being compared and analysed:

PriS *"Privacy Safeguard"* a privacy requirements gathering system (Kavakli et al. 2006) that uses privacy requirements as organisational goals to incorporate privacy into processes and systems (Kalloniatis et al. 2008);

IRIS/CAIRIS the IRIS *"Integrating Requirements and Information Security"* (Faily and Fléchais 2009) and CAIRIS *"Computer Aided Integration of Requirements and Information Security"* (Faily and Fléchais 2010) method seek to elicit and visualise security requirements, vulnerabilities and threats using goal modelling. This framework now also incorporates privacy considerations (Faily 2018);

LINDDUN A privacy privacy threat modelling framework (Deng et al. 2011). LINDDUN stands for: Linkability, Identifiability, Non-repudiation, Detectability, Disclosure, content Unawareness and policy and consent Non-compliance;

Pfitzmann & Hansen A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management (Pfitzmann and Hansen 2010);

Schleswig-Holstein DPA Studies by the Schleswig-Holstein Data Protection Authority, the Independent Centre for Privacy Protection (ICPPS) around privacy terminology and goals: (Hansen 2012, Hansen et al. 2015);

ENISA Guidance from the European Union Agency for Network and Information Security (ENISA) around data protection and privacy goals (Danezis et al. 2014); and

ISO the ISO guides for security and privacy risk (BS ISO 27000:2017 2017, BS ISO/IEC 29100 2011).

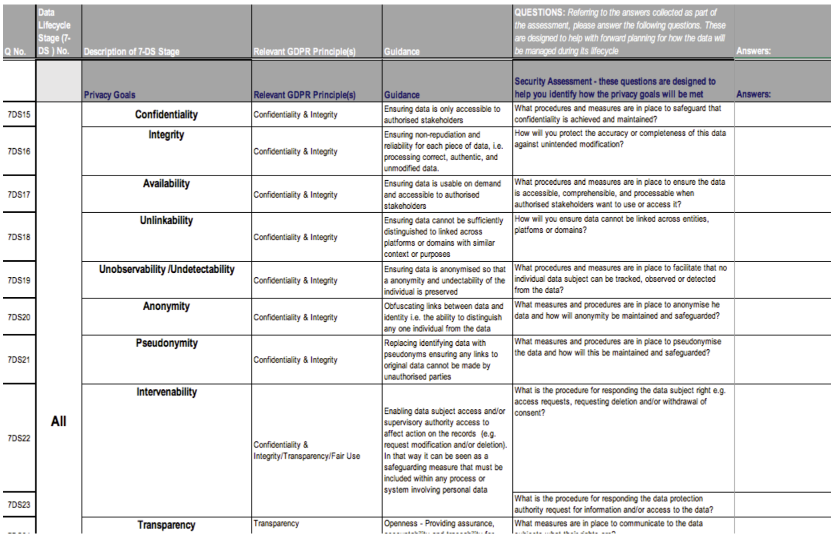
The analysis itself took the form of first, looking at what each framework considered a privacy goal and how each goal was defined. To analyse frameworks a spreadsheet was used to compare the goals. This resulted in 16 goals being identified between the 6 frameworks, these can be found in Table 8.2.

No.	Goal	Mentioned in following Frameworks	Privacy Meaning	PLAN
1.	Confidentiality	LINDDUN, IRIS (CAIRIS), ICPPS, ENISA - ISO mention but not as a goal	*Ensuring data is usable on demand and accessible to authorised stakeholders.*	Part of the CIA triangle (Confidentiality, Integrity and Availability), widely accepted as standard practice to include for both security and privacy (Danezis et al. 2014). Included in PLAN

2.	Integrity	LINDDUN, IRIS (CAIRIS), ICPPS - ENISA, PriS - ISO mention but not as a goal	*Ensuring non-repudiation and reliability for each piece of data, i.e. processing correct, authentic, and unmodified data.*	Part of the CIA triangle. Included in PLAN
3.	Availability	LINDDUN, IRIS (CAIRIS), ICPPS, ENISA - ISO mention but not as a goal	*Ensuring data is usable on demand and accessible to authorised stakeholders.*	Part of the CIA triangle. Included in PLAN
4.	Unlinkability	PriS, IRIS (CAIRIS), Pfitzmann & Hansen 2010, ENISA, ICPPS	*Ensuring data cannot be sufficiently distinguished to linked across platforms or domains with similar context or purpose.*	Included in PLAN
5.	Unobservability	PriS, IRIS (CAIRIS), Pfitzmann & Hansen 2010- ICPPS mentions but not as a goal	*Ensuring no unauthorised party can observe what data or service is being utilised or performed, even if they gain access to the system.*	Included in PLAN with *Undetectability* (see Line 11). In terms of data, both are mechanisms for hiding information to facilitate the unobservability and undetectability of the data concerning the individual data subject. It could be argued these should be split into two separate goals. The reason these two have been grouped in PLAN is because, unobservability refers to whether the users actions (e.g. sending or receiving) are observable and ensuring that the data itself cannot be observed either. This, arguably includes undetectability as, if an attacker or unauthorised person, cannot observe then they cannot detect, thus, in PLAN, these terms have been grouped.
6.	Anonymity	LINDDUN, PriS, IRIS (CAIRIS), Pfitzmann & Hansen 2010, ENISA - ICPPS and ISO mention but not as a goal	*Obfuscating links between data and identity i.e. the ability to distinguish any one individual from the data.*	Included in PLAN
7.	Pseudonymity	LINDDUN, PriS, IRIS (CAIRIS), Pfitzmann & Hansen 2010- ENISA, ICPPS and ISO mention but not as a goal	*Replacing identifying data with pseudonyms ensuring any links to original data cannot be made by unauthorised parties.*	Included in PLAN
8.	Intervenability	ENISA 2014, ICPPS	*Enabling data subject access and/or supervisory authority access to affect action on the records (e.g. request modification and/or deletion). In that way it can be seen as a safeguarding measure that must be included within any process or system involving personal data.*	Requirement for compliance with GDPR and therefore included in PLAN
9.	Transparency	ENISA 2014, ICPPS- PriS & ISO mention but not as a goal	*Openness - Providing assurance, accountability and traceability for internal and external stakeholders.*	Requirement for compliance with GDPR and therefore included in PLAN
10.	Authentication	LINDDUN, PriS - Pfitzmann & Hansen, ICPPS & ENISA mention but not as a goal		Only LINDDUN and PriS consider this a privacy goal. Arguably this is a system or process setting requirement rather than a goal - it is a validation that the person/system is who they claim to be which, in privacy terms, should be covered as part of integrity. Therefore this has not been included as a privacy goal in PLAN
11.	Undetectability	Pfitzmann & Hansen 2010- ICPPS mention but not as a goal	*Ensuring data is annonymised so that anonymity and undectability of the individual is preserved.* & *Ensuring data cannot be sufficiently distinguished to establish whether it exists or not*	Included in PLAN together with *unobservability* (see Line 5).
12.	Accountability	LINDDUN, IRIS (CAIRIS), ENISA, Pfitzmann & Hansen, ICPPS & ISO29100 mention but not as a goal		Arguably this relates to compliance and transparency as, in order to meet compliance, there has to be accountability. In privacy terms this can be related to having an accountable person and having accountability over what data is collected, used, processed or shared. Therefore, in PLAN this is considered to be incorporated within transparency.
13.	Identification	PriS		Not included in PLAN. This concerns user identifying themselves in a system as opposed to identification of an individual or risk of re-identification which is the opposite of identifiability and therefore not a privacy goal.
14.	Data Protection	PriS		This is considered embedded within all of the above privacy goals. Not included in PLAN.
15.	Authorisation	LINDDUN, PriS, IRIS (CAIRIS) mention but not as goal		This is similar to authentication in that a system or process will need to have this property for validation purposes but, for privacy, this forms part of integrity and confidentiality. Not included in PLAN.
16.	Non-repudiation	LINDDUN, ENISA mentions this as a threat, not a privacy goal - ICPPS & IRIS (CAIRIS) mention but not as goal		As above, this is part of integrity and thus, not included in PLAN.

Table 8.2: Privacy Goal Comparison

Table 8.2 shows that the first 7 goals (no's 1-7) should be classed as privacy goals and therefore these were included in PLAN. The next two goals (no's 8 and 9), were classed as goals by some researchers, but others did not mention them. However, as these are requirements for GDPR, these too were included as privacy goals to be included in PLAN and incorporated into the questions. This resulted in 9 privacy goals being added to the PLAN, these are depicted in Figure 8.8. The full list of questions and guidance provided in the PLAN can be found in Appendix F, Section F.6.1.



Q No.	Data Lifecycle Stage (7-DS) No.	Description of 7-DS Stage	Relevant GDPR Principle(s)	Guidance	QUESTIONS: *Referring to the answers collected as part of the assessment, please answer the following questions. These are designed to help with forward planning for how the data will be managed during its lifecycle*	Answers:
		Privacy Goals	Relevant GDPR Principle(s)	Guidance	Security Assessment - these questions are designed to help you identify how the privacy goals will be met	Answers:
7DS15	All	**Confidentiality**	Confidentiality & Integrity	Ensuring data is only accessible to authorised stakeholders	What procedures and measures are in place to safeguard that confidentiality is achieved and maintained?	
7DS16		**Integrity**	Confidentiality & Integrity	Ensuring non-repudiation and reliability for each piece of data, i.e. processing correct, authentic, and unmodified data.	How will you protect the accuracy or completeness of this data against unintended modification?	
7DS17		**Availability**	Confidentiality & Integrity	Ensuring data is usable on demand and accessible to authorised stakeholders	What procedures and measures are in place to ensure the data is accessible, comprehensible, and processable when authorised stakeholders want to use or access it?	
7DS18		**Unlinkability**	Confidentiality & Integrity	Ensuring data cannot be sufficiently distinguished to linked across platforms or domains with similar context or purposes	How will you ensure data cannot be linked across entities, platfoms or domains?	
7DS19		**Unobservability /Undetectability**	Confidentiality & Integrity	Ensuring data is anonymised so that a anonymity and undectability of the individual is preserved	What procedures and measures are in place to facilitate that no individual data subject can be tracked, observed or detected from the data?	
7DS20		**Anonymity**	Confidentiality & Integrity	Obfuscating links between data and identity i.e. the ability to distinguish any one individual from the data	What measures and procedures are in place to anonymise he data and how will anonymity be maintained and safeguarded?	
7DS21		**Pseudonymity**	Confidentiality & Integrity	Replacing identifying data with pseudonyms ensuring any links to original data cannot be made by unauthorised parties	What measures and procedures are in place to pseudonymise the data and how will this be maintained and safeguarded?	
7DS22		**Intervenability**	Confidentiality & Integrity/Transparency/Fair Use	Enabling data subject access and/or supervisory authority access to affect action on the records (e.g. request modification and/or deletion). In that way it can be seen as a safeguarding measure that must be included within any process or system involving personal data	What is the procedure for responding the data subject right e.g. access requests, requesting deletion and/or withdrawal of consent?	
7DS23					What is the procedure for responding the data protection authority request for information and/or access to the data?	
		Transparency	Transparency	Openness - Providing assurance,	What measures are in place to communicate to the data	

Figure 8.8: PLAN list of privacy goals

The PLAN replaces the decision phase in CLIFOD and the Consultation stage in the DPIA Data Wheel. Here, rather than make a decision, practitioners are asked to make a plan for how they intend to manage the data during its time with the organisation, which involves not one, but many decisions relating to the data.

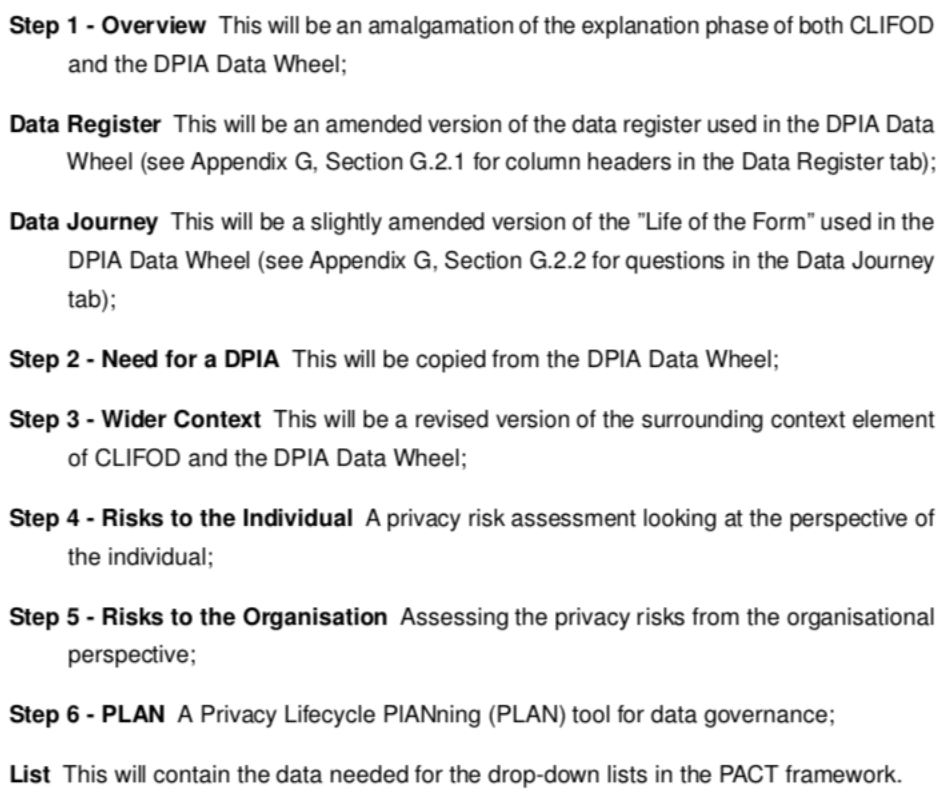
Once practitioners have completed the PLAN, the assessment is complete.

8.5.4 PACT Manner of Presentation

The PACT framework will be a paper prototype (see Chapter 3, Section 3.4.11), captured in a spreadsheet that will consist of 10 worksheets ('tabs'), a list of these can be found in Figure 8.9.

8.5.5 Evaluating the Risk Assessment and PLAN

Because the PLAN is 'new' in terms of this work, this was separately evaluated along with the revised risk assessments and format. This evaluation involved sending the PACT prototype spread sheet to three groups of evaluators for feedback on the PLAN and the guidance and scoring included in the risk assessments. The aim of this evaluation was to get a broad range of actual and potential 'practitioners' to review these sections of the framework as the intention is that the final framework should work for both any practitioner,



Step 1 - Overview This will be an amalgamation of the explanation phase of both CLIFOD and the DPIA Data Wheel;

Data Register This will be an amended version of the data register used in the DPIA Data Wheel (see Appendix G, Section G.2.1 for column headers in the Data Register tab);

Data Journey This will be a slightly amended version of the "Life of the Form" used in the DPIA Data Wheel (see Appendix G, Section G.2.2 for questions in the Data Journey tab);

Step 2 - Need for a DPIA This will be copied from the DPIA Data Wheel;

Step 3 - Wider Context This will be a revised version of the surrounding context element of CLIFOD and the DPIA Data Wheel;

Step 4 - Risks to the Individual A privacy risk assessment looking at the perspective of the individual;

Step 5 - Risks to the Organisation Assessing the privacy risks from the organisational perspective;

Step 6 - PLAN A Privacy Lifecycle PlANning (PLAN) tool for data governance;

List This will contain the data needed for the drop-down lists in the PACT framework.

Figure 8.9: PLAN list of Worksheets ('Tabs')

irrespective of background and experience. It should not have to rely on expert input as not all organisations will have such expertise to hand. Therefore, three groups of initial reviewers were selected to capture differing levels of expertise and perspective as follows:

Academics The spreadsheet was sent to three academics for review; one with expertise in IT and computing; one with a security background; and one with a methodology and requirements engineering background;

Practitioners Three practitioners reviewed the framework; a Data Protection Officer (DPO); an IT Security Manager; and a Consultant with a management and IT project implementation background;

Peers The last group of evaluators were three fellow PhD students who study risk, security and computing.

From this, the reviewers liked the new format for the risk assessments and the explanations provided were considered *"informative and conducive to understanding how this might work"* (R5), while another commented on the choice of using the CNIL framework asking why that particular framework had been chosen for the scoring. The answer provided was *as far as [the researcher] has been able to ascertain, this is the only framework that specifically looks at how to score risks to the individual rather than the organisation* *and, it does so rather well*. As regards the PLAN, one asked why those particular privacy goals were chosen (referred the reviewer to Table F.5 and explanation above) and one responded with some minor rephrasing of questions suggestions, commenting: *"... it looks good and captures everything that we, the InfoSec team, would look for. I think seeing it "used in anger" will be a really useful exercise and I can't wait to see the outcomes"* (R3). The other reviews mainly consisted of minor corrections and suggestions for amendments, all of which were incorporated into the framework. These amendments are included in the description of the questions and layout as outlined in the sections above and Appendix F.

8.6 Stage 3 - Evaluating PACT

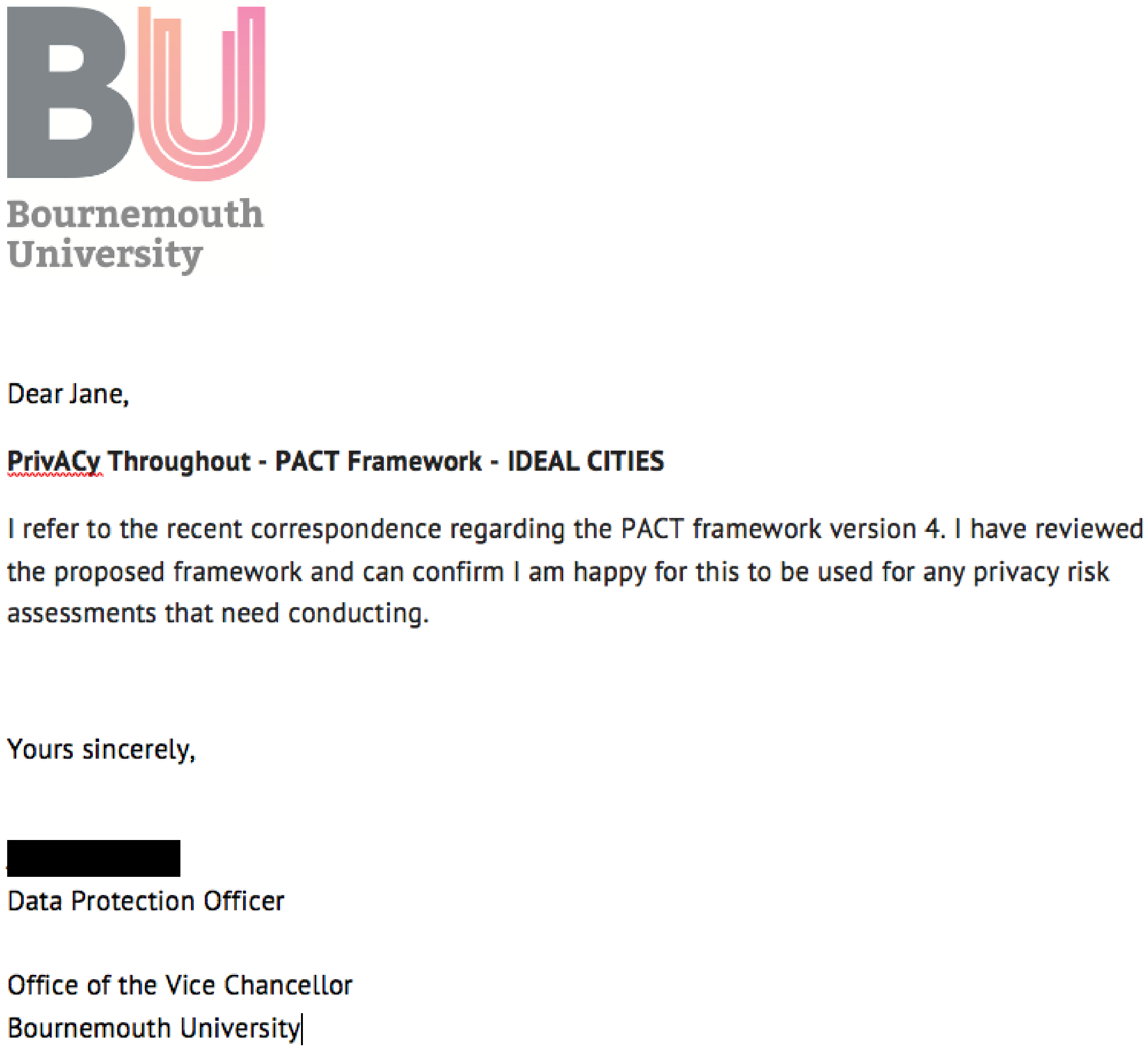
For the evaluation of the whole of the PACT framework, the researcher travelled to Cyprus and worked for one month with one of the industry partners on the Ideal-Cities project, Cablenet (CBN). During this time, the initial PACT framework was devised as set out in the sections above and a preliminary evaluations was conducted before completing a full evaluation with CBN, the intention being that the PACT framework will form part of the ethics and privacy compliance requirements for the whole Ideal-Cities project, placing an obligation on all the project partners to complete the PACT framework.

Further, because the intention was for this framework to become an integral and compulsory framework to be used for a European Horizon 2020 project (*Ideal-Cities*), the DPO who evaluated the PLAN, also agreed to evaluate the complete PACT framework and, as a result, endorsed the framework for use *"for any privacy risk assessment that needs conducting"*. A copy of the endorsement letter can be found in Figure 8.10.

8.6.1 Evaluating 1 - Ideal-Cities

For the second part of the evaluation, the idea for this work was to collaborate with CBN on devising an ethics and privacy framework for the whole Ideal Cities project and, as part of this work, evaluate the PACT framework with CBN with a view to making completing the PACT framework a compulsory part of the Ideal Cities project. This evaluation would consist of a more in-depth evaluation and completion of the whole of the PACT framework in line with the protocol. However, due to time constraints and availability of the collaborating practitioner, there was not enough time to complete the full two evaluations originally planned for in the protocol (see Appendix 8.2, Section 8.2.3). Instead only part of the first evaluation was conducted, assessing a new project, the Ideal-Cities project from CBN's perspective. This evaluation resulted in 22 attributes being identified, 12 personal identifiers (*PIs*) and 8 quasi-identifiers (*QI*, see Chapter 2, Section 2.4.3) and the overview being completed.

The evaluation took approximately 4 hours to complete over a series of 4 meetings



Bournemouth University

Dear Jane,

PrivACy Throughout - PACT Framework - IDEAL CITIES

I refer to the recent correspondence regarding the PACT framework version 4. I have reviewed the proposed framework and can confirm I am happy for this to be used for any privacy risk assessments that need conducting.

Yours sincerely,

[redacted]

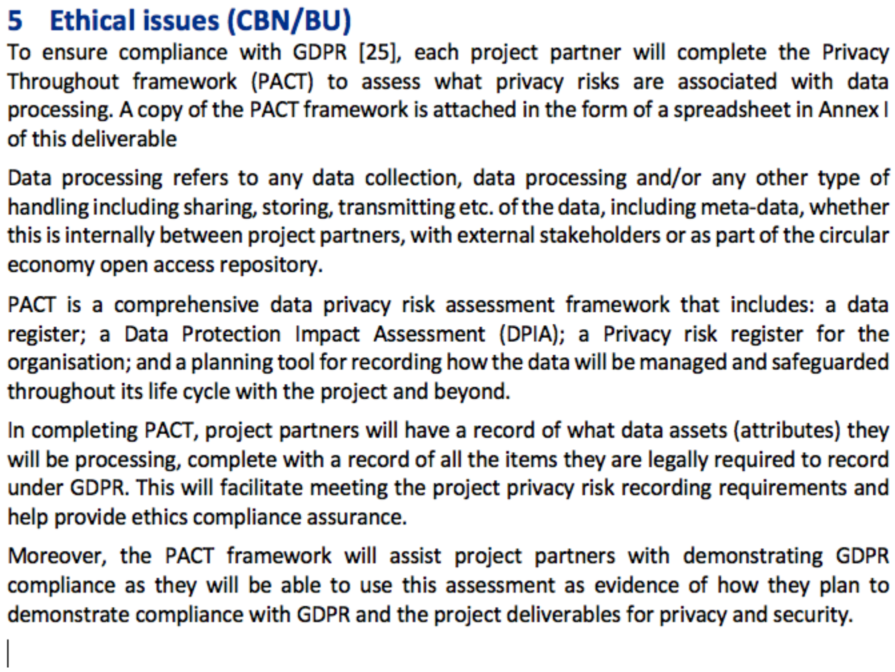
Data Protection Officer

Office of the Vice Chancellor
Bournemouth University

Figure 8.10: DPO letter of Endorsement

at CBN's offices. From this it transpired that, as it was, the CBN practitioner felt, first that overall, the framework as a whole might be *"a bit long"* and difficult to complete without guidance, particularly for a non-privacy or security practitioner. To help address this, it was agreed as part of the meetings that various aspects of the PACT framework should perhaps be rearranged. For example, placing the data register and data journey at the beginning of the framework or even completely separating these out to allow for multiple stakeholders to be given an opportunity to complete these forms which, after all, are meant to be used to inform the rest of the framework. Further, the CBN practitioner suggested that perhaps all of the worksheets should be presented as separate spreadsheets so as to break the framework down into more manageable 'bitesize' chunks as perhaps this could potentially help facilitate both part of the consultation element and provide the DPO at the organisation with an opportunity to review these before the risk assessment(s) are completed. Moreover, for the 'Risks to the Individual' assessment, it was suggested that perhaps this could be renamed 'Risks to the Data Subject' to align with GDPR terminology. In addition, a number of the questions, particularly the context questions, would need to be revised or reworded to clarify what was required.

Therefore, it was agreed a workshop would be arranged at a later data after the evaluations were complete, where all six project partners working on the project (see Section 8.1) would be given an opportunity to work with the researcher and the CBN practitioner in completing the PACT spreadsheet for their planned data collection for the Ideal-Cities project. This workshop is planned for early March 2019. However, the intention was for the PACT framework to be incorporated into the Ideal-Cities project to become the privacy assessment framework for the whole project. The deadline for for uploading the project Data Management Plan (DMP) to the EU Commission was set for 31st December 2018 (see Figure 8.1). Thus, in order to complete the evaluations and the framework in time for this deadline and be ready for the planned workshops the evaluations would need completing prior to this deadline (see Figure 8.11).



5 Ethical issues (CBN/BU)

To ensure compliance with GDPR [25], each project partner will complete the Privacy Throughout framework (PACT) to assess what privacy risks are associated with data processing. A copy of the PACT framework is attached in the form of a spreadsheet in Annex I of this deliverable

Data processing refers to any data collection, data processing and/or any other type of handling including sharing, storing, transmitting etc. of the data, including meta-data, whether this is internally between project partners, with external stakeholders or as part of the circular economy open access repository.

PACT is a comprehensive data privacy risk assessment framework that includes: a data register; a Data Protection Impact Assessment (DPIA); a Privacy risk register for the organisation; and a planning tool for recording how the data will be managed and safeguarded throughout its life cycle with the project and beyond.

In completing PACT, project partners will have a record of what data assets (attributes) they will be processing, complete with a record of all the items they are legally required to record under GDPR. This will facilitate meeting the project privacy risk recording requirements and help provide ethics compliance assurance.

Moreover, the PACT framework will assist project partners with demonstrating GDPR compliance as they will be able to use this assessment as evidence of how they plan to demonstrate compliance with GDPR and the project deliverables for privacy and security.

Figure 8.11: Ideal Cities Data Management Plan excerpt

Thus, to complete the evaluation of PACT, two further evaluations were arranged. For time and logistical reasons it made sense to try to conduct these evaluation with locally based industry organisations rather than Ideal-Cities project partners. Further, to seek to triangulate the evaluations (Yin 2013) of the PACT framework, it would make sense if PACT could be evaluated by practitioners from three different industry sectors: the public sector, the private sector (CBN) and the charity sector (UKNC). Therefore two local business were approached: a UK public body organisation (PBO); and a UK national Charity (UKNC) with offices based locally. Both agreed to take part and help evaluation the PACT framework.

8.6.2 Evaluating 2 - Public Body

The second evaluation took place with the PBO at their offices over two sessions, one lasting 3 hours, the second 1.5 hours. This evaluation reviewed the PBO's existing process for dealing with security incidents using an off-the-shelf incident management software package. The PACT framework was completed considering the process for managing and recording an incident within this system.

In completing the data register, it was determined there are 10 different attributes collected and recorded, 7 of which were classed as PIs (personal identifiers) and 3 as QI (quasi-identifiers) attributes. These attributes go on two journeys as part of their processing, after which the records are stored indefinitely according to current practice.

The evaluation then continued by completing each of the 6 steps of PACT.

PBO - Step 1

Completing the overview (Step 1) confirmed the indefinite storage period to be standard current practice and also revealed that the data may be shared with external stakeholders such as the ICO as part of reporting although, this was deemed that would be acceptable in the circumstances, as this is a legal requirement if a breach has occurred (GDPR, Article 33, see Chapter 2, Section 2.4.4). However, for the indefinite storage period, the practitioner acknowledged this is contrary to a number of GDPR principles including storage limitation (*principle 5*), data minimisation (*principle 3*), lawfulness, fairness and transparency (*principle 1*) and, as such, could pose a threat to data integrity and confidentiality (*principle 6*) and breach the rights of the data subject. This was reflected in both the risk assessments and noted as a risk to both the data subject (in the 'risks to the individuals' section) and to the organisation (in the 'risks to the organisation' section). Therefore, addressing this was noted as an action be be implemented urgently in the mitigation strategy section and also noted within the PLAN (under *unobservability/undetectability, transparency and storage limitation*, Step 6).

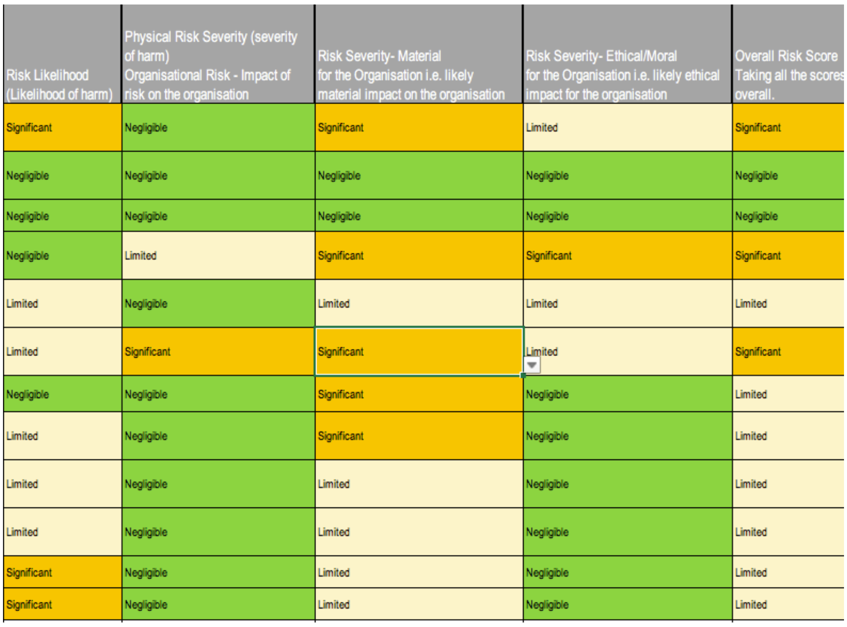
On the overview, the idea of recording relationships between actors caused some confusion, the practitioner was not quite sure what this meant. Once explained, the questions were answered with no difficulties but this did highlight that perhaps this needed more explanation. Moreover, the practitioner pointed out that repeating the same question for each actor seemed unnecessary, suggesting to instead ask the question once for all actors.

PBO - Step 2

The 'Need for a DPIA' (Step 2) assessment confirmed the need for a DPIA (both required and advisable). The practitioner commented that there may be multiple legal basis upon which the data is processed and therefore perhaps multiple answers should be selectable from the drop-down list.

PBO - Steps 3 & 4

Both the 'risks to the individual' (Step 3) and the 'risks to the organisation' (Step 4) were completed. This resulted in 3 risks to the individual being identified, with one risk receiving a score of *significant* for the likelihood of harm where personal data is leaked to an external third-party. All other scores on this risk assessment were scored and scored *negligible* or *limited*. For the risks to the organisation, 11 risks were noted, these included the 3 risks to the individual identified in the risks to the individual's risk assessment and a number of more specific or generic security risks that could potentially result in privacy being compromised such as *"malware being installed on [PBO] system or device"*; *"intentional insider threat - e.g disgruntled staff"*; and *"unintentional insider threat - e.g untrained staff"*. Most risks to the organisation were scored *negligible* or *limited*, there were however, a number of *significant* ratings noted as well (see Figure 8.12).



Risk Likelihood (Likelihood of harm)	Physical Risk Severity (severity of harm) Organisational Risk - Impact of risk on the organisation	Risk Severity- Material for the Organisation i.e. likely material impact on the organisation	Risk Severity- Ethical/Moral for the Organisation i.e. likely ethical impact for the organisation	Overall Risk Score Taking all the scores overall.
Significant	Negligible	Significant	Limited	Significant
Negligible	Negligible	Negligible	Negligible	Negligible
Negligible	Negligible	Negligible	Negligible	Negligible
Negligible	Limited	Significant	Significant	Significant
Limited	Negligible	Limited	Limited	Limited
Limited	Significant	Significant	Limited	Significant
Negligible	Negligible	Significant	Negligible	Limited
Limited	Negligible	Significant	Negligible	Limited
Limited	Negligible	Limited	Negligible	Limited
Limited	Negligible	Limited	Negligible	Limited
Significant	Negligible	Limited	Negligible	Limited
Significant	Negligible	Limited	Negligible	Limited

Figure 8.12: Risk to the Organisation - Scores PBO

PBO - Step 5

In Step 5, the wider context, this section invoked quite a lot of discussion as, initially, the practitioner felt that none of the categories were relevant. However, after some discussion, the practitioner did come up with a number of risks to be noted including, for potential infringement of social values; *”If access to certain sites, social media or systems were blocked in response to an incident, this could be perceived as an infringement of social norms e.g. a staff member (data subject) might expect to be able to connect via social media from a [PBO] device”*. Thus, it appeared that, once the idea of relating the context to the process being assessed had been explained, it appeared that context *does* have an impact on data processing and how this might be perceived. However, it may be that these discussions should take place earlier, before the risk assessments are completed. Also, looking at the process of completing PACT from a more practical framework flow perspective, the responses entered in both the column for the data subject perspective and the organisational perspectives, were more or less identical, meaning that perhaps having just one answer column here would suffice.

PBO - Step 6

This step, according to the practitioner had *”a number of similar features”* to the type of planning the PBO does currently, but with more *”emphasis on GDPR and direction for what should be considered”*. Some of the questions needed input from other departments to answer in detail and for these, a note was put on the PLAN worksheet to the effect that *”handed over to [relevant] team”* with a view that, after the session the practitioner would liaise with these teams and then update with fuller answers. In view of this, perhaps a note needs to be added to the effect that the PLAN should be completed in collaboration with all stakeholders involved. The practitioner also mentioned that it would be useful if some of the aspects, such as the legal basis for processing, could be made to auto-complete from previous answers provided.

8.6.3 Evaluation 3 - UK National Charity

The second external industry evaluation with a UK National Charity took the format of a meeting with the Data Protection Officer (DPO) to discuss the PACT framework. As part of this discussion a number or issues/concerns were raised by the practitioner. First, splitting the context into two sections would, according to this practitioner *”just encourage me to skip the context section altogether”*. Second, the practitioner pointed out that what normally happens within their organisation is that non-privacy experts are asked to complete a privacy impact assessment (PIA) for any new project. This PIA is short and, while the practitioner acknowledged, this will need replacing with something more comprehensive, the PACT framework, as it is currently, would scare most staff: *"it is hard enough getting staff to complete our current PIA (which in comparison to PACT is "short and sweet"), if I sent them this they would just get scared and not bother at all"*. Therefore, a discussion ensued about how the PACT framework might be broken down and given to staff in *stages*. For example, splitting the spreadsheet into multiple spreadsheets, sending one worksheet's worth at a time (so to speak) to staff (e.g. send the data register first, then the data journey etc.). That way, once complete, depending on the responses received, the next spreadsheet (worksheet) can be forwarded and, where necessary, support provided as needed to fill in individual sections or worksheets. Alternatively, if the framework was presented as an automated tool, this would, of course, resolve a lot of these issues as the tool would move from one question to the next without presenting the whole framework in one go.

One area that this practitioner felt particularly strongly needed addressing was the lack of: *"ability to quickly and/or easily demonstrate compliance, there is currently no one place where you can go to look to find out whether you meet all the criteria for completing the DPIA if the ICO [the authorities] come to check, that is a major flaw in my opinion. At the very least, everything relating to the DPIA should be on the same worksheet"*.

The practitioner also commented that they felt the categorisation for the attributes on the data register was useful, particularly how this had been divided into sections depending on what was being assessed. Further, they felt that splitting the risks into sections that covered the different perspectives and the accompanying guidance and instructions was *"really helpful"*.

8.7 Post-evaluation changes to PACT

Based on the evaluations and the feedback provided by the practitioners a number of changes were made to the PACT spreadsheet and the order of the steps were moved and merged as follows:

- The Data Register and Data Journey have been moved to the beginning of the framework, to be completed before the assessment begins. This will allow for these worksheets to be separated out and (i) sent to multiple stakeholders for completion if required and (ii) enable some pre-assessment to be conducted by someone with appropriate privacy, security and/or legal knowledge. This way, the pre-assessor can determine first, who needs to be involved in completing the assessment and second, what level of support and/or guidance is needed for the assessment. On the data register, some of the headings were reworded slightly as follows;

  Attribute Name This refers to the individual data element i.e. the Asset or Data Asset - Where a lot of data is processed, this may refer to a group of assets that fit into a particular attribute or asset category;

Storage Method please note how the data will be stored;

Data Security provide a brief description for how the data will be safeguarded (you may wish to complete this after the assessment(s) are complete);

legal/lawful basis for processing *Primary* legal/lawful basis for processing: select from drop-down.

- *Step 1 'Overview'* was moved to become Step 2, a number of additional changes were made to this step which are described in more detail in Section 8.7.1;

- *Step 2 'Need for a DPIA'* was moved to the beginning of the risk assessment to become 'Step 1' to facilitate the observation that the answers here could influence how the overview and risks assessments will be completed. Furthermore, additional rows were added to each of the questions relating to lawfulness of processing (see Appendix F, Section F.3);

- *Step 3 'Wider Context'* was amalgamated into the overview section (see Section 8.7.1);

- *Step 4 'Risks to the Individual'* was renamed 'Risks to the Data Subject' and moved to become Step 3. In the 'residual risk' section, a column was added for *'Residual Risk (DS) Severity score post-mitigation'* as this had been omitted in error on the PACT spreadsheet;

- *Step 5 'Risks to the Organisation'* was moved to become Step 4. Moreover, in the 'residual risk' section, a column was added for *'Residual Risk (Org) Severity score post-mitigation'* as this had been omitted in error on the PACT spreadsheet;

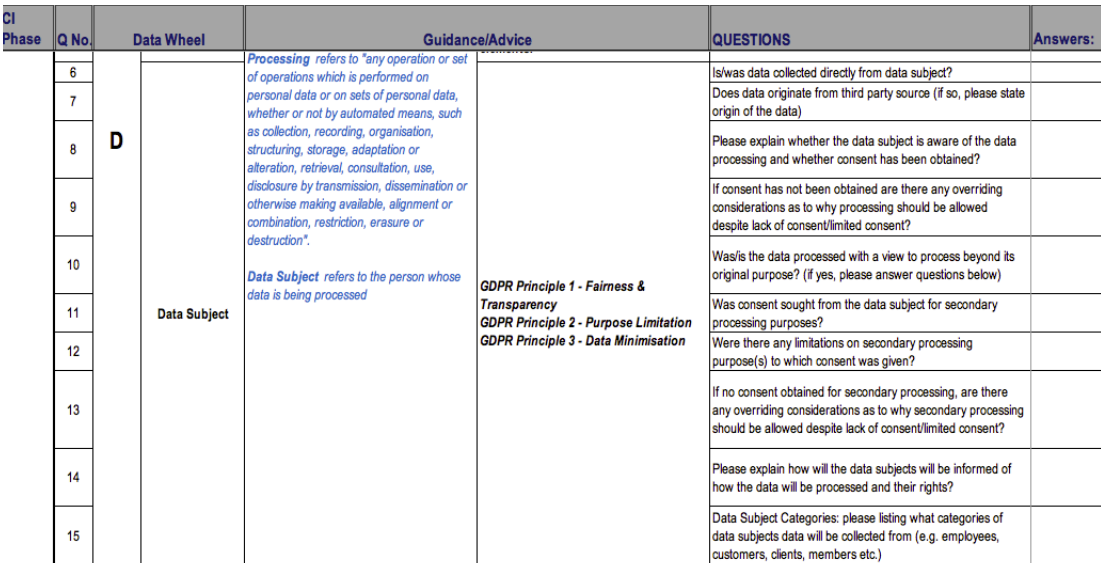
- *Step 6 'PLAN'* was moved to become Step 5. The answer columns were amalgamated into one column that now reads *'Answers'* only rather than split this into perspectives and a paragraph has been added to the instructions at the top of the worksheet to read: *It is recommended that you complete this section in collaboration with all relevant stakeholders involved to ensure a comprehensive data management plan is in place. Once complete you may wish to also consult Management and your Privacy Officer an/or Data Protection Officer (DPO) for a review of the PLAN.*

To reflect these moves, the instructions and headers for each step were updated so that the order of the steps referred to on each worksheet was in alignment with the changes made and the questions in the various worksheets were renumbered to account for the changes made in the overview (see Section 8.7.1), these changes are described in more detail in the next section.

8.7.1 Overview

The overview is where most of the changes were made. These included slightly rewording some of the questions to make the meaning a bit clearer, moving some questions (see below) and renumbering the questions and adding a column to show which GDPR Principle the question relates to.

Actors/Relationships



CI Phase	Q No.	Data Wheel		Guidance/Advice		QUESTIONS	Answers:
	6	D	Data Subject	***Processing*** *refers to "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction".* ***Data Subject*** *refers to the person whose data is being processed*	***GDPR Principle 1 - Fairness & Transparency*** ***GDPR Principle 2 - Purpose Limitation*** ***GDPR Principle 3 - Data Minimisation***	Is/was data collected directly from data subject?	
	7					Does data originate from third party source (if so, please state origin of the data)	
	8					Please explain whether the data subject is aware of the data processing and whether consent has been obtained?	
	9					If consent has not been obtained are there any overriding considerations as to why processing should be allowed despite lack of consent/limited consent?	
	10					Was/is the data processed with a view to process beyond its original purpose? (if yes, please answer questions below)	
	11					Was consent sought from the data subject for secondary processing purposes?	
	12					Were there any limitations on secondary processing purpose(s) to which consent was given?	
	13					If no consent obtained for secondary processing, are there any overriding considerations as to why secondary processing should be allowed despite lack of consent/limited consent?	
	14					Please explain how will the data subjects will be informed of how the data will be processed and their rights?	
	15					Data Subject Categories: please listing what categories of data subjects data will be collected from (e.g. employees, customers, clients, members etc.)	

Figure 8.13: Overview - Data Subject

First, in the *"Actors"* section all questions relating to the Data Subject were moved to a new section at the beginning of the overview entitled *"Data Subject"*, see Figure 8.13. Here, all questions from the actors section (other than the relationship reference noted below) were placed along with the questions from the prevailing context section that referred to consent (Q36-45, see Appendix F, Section F.4).

Another change made in this section involved removing the questions about relationships for each actor group (DPO, DC etc.) and merging these into one section to cover all relationships that asks:

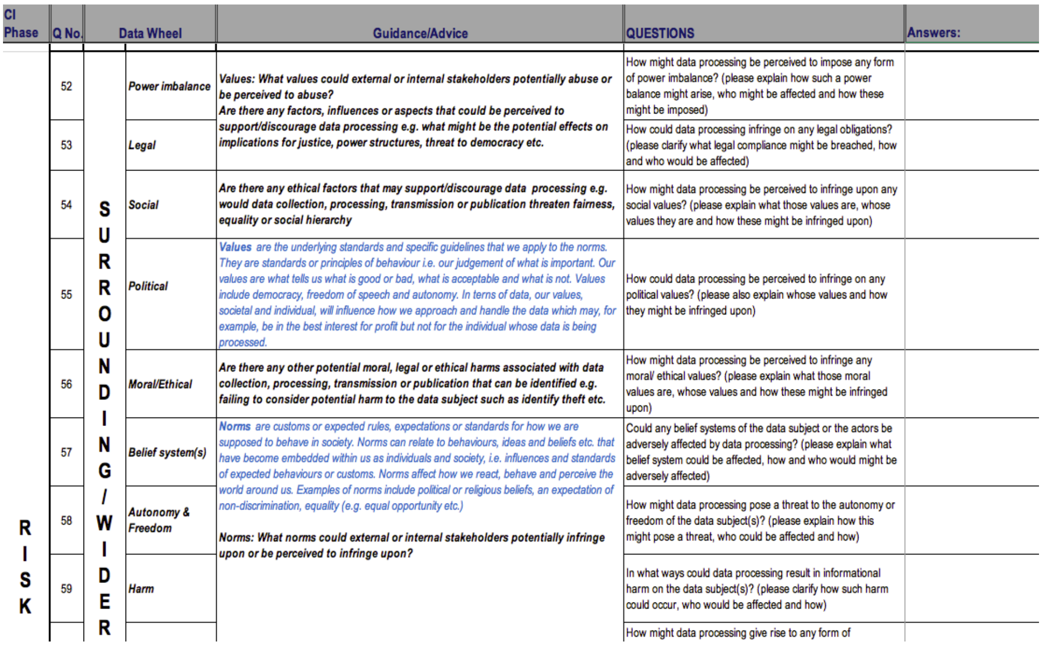
Q47 What is the relationship between the data subject and other actors? (e.g. friends, co-worker, professionally, citizen, employment) i.e. how do they interact with each other. Please note all relationships that apply; and

Q48 What is the relationship between processing actors? (e.g. friends, co-worker, professionally, citizen, employment) i.e. how do they interact with each other, for example, it may be that the same actors are both friends and employer/Manager or employee. Similarly, the data controller and the data processor could have both a professional relationship (supplier/customer) and be friends or colleagues. Please note all relationships that apply.

Attributes

The attributes section was expanded upon with some additional questions to ensure all information needed to demonstrate that the DPIA meets GDPR requirements have been met (GDPR, Article 35, ICO guidance (see Chapter 2, Section 2.4.4) can be found on the overview worksheet. This was done so that, if for example, a practitioner decides to split the framework into multiple spreadsheets, as suggested by a couple of the practitioners in the evaluations, everything needed to demonstrate compliance can be found on the one worksheet. This does mean that some of the questions might be repeated on the PLAN but this can be easily be addressed and the relevant fields made to pre-populate, if (when) the prototype spreadsheet is turned into a prototype computerised tool.

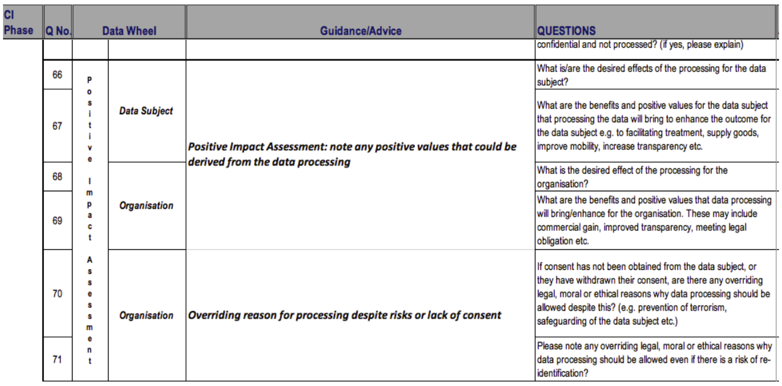
Context



CI Phase	Q No.	Data Wheel		Guidance/Advice	QUESTIONS	Answers:
RISK	52	SURROUNDING/WIDER	*Power imbalance*	***Values: What values could external or internal stakeholders potentially abuse or be perceived to abuse?*** ***Are there any factors, influences or aspects that could be perceived to support/discourage data processing e.g. what might be the potential effects on implications for justice, power structures, threat to democracy etc.***	How might data processing be perceived to impose any form of power imbalance? (please explain how such a power balance might arise, who might be affected and how these might be imposed)	
	53		*Legal*		How could data processing infringe on any legal obligations? (please clarify what legal compliance might be breached, how and who would be affected)	
	54		*Social*	***Are there any ethical factors that may support/discourage data processing e.g. would data collection, processing, transmission or publication threaten fairness, equality or social hierarchy***	How might data processing be perceived to infringe upon any social values? (please explain what those values are, whose values they are and how these might be infringed upon)	
	55		*Political*	***Values*** *are the underlying standards and specific guidelines that we apply to the norms. They are standards or principles of behaviour i.e. our judgement of what is important. Our values are what tells us what is good or bad, what is acceptable and what is not. Values include democracy, freedom of speech and autonomy. In terns of data, our values, societal and individual, will influence how we approach and handle the data which may, for example, be in the best interest for profit but not for the individual whose data is being processed.*	How could data processing be perceived to infringe on any political values? (please also explain whose values and how they might be infringed upon)	
	56		*Moral/Ethical*	***Are there any other potential moral, legal or ethical harms associated with data collection, processing, transmission or publication that can be identified e.g. failing to consider potential harm to the data subject such as identify theft etc.***	How might data processing be perceived to infringe any moral/ ethical values? (please explain what those moral values are, whose values and how these might be infringed upon)	
	57		*Belief system(s)*	***Norms*** *are customs or expected rules, expectations or standards for how we are supposed to behave in society. Norms can relate to behaviours, ideas and beliefs etc. that have become embedded within us as individuals and society, i.e. influences and standards of expected behaviours or customs. Norms affect how we react, behave and perceive the world around us. Examples of norms include political or religious beliefs, an expectation of non-discrimination, equality (e.g. equal opportunity etc.)* ***Norms: What norms could external or internal stakeholders potentially infringe upon or be perceived to infringe upon?***	Could any belief systems of the data subject or the actors be adversely affected by data processing? (please explain what belief system could be affected, how and who would might be adversely affected)	
	58		*Autonomy & Freedom*		How might data processing pose a threat to the autonomy or freedom of the data subject(s)? (please explain how this might pose a threat, who could be affected and how)	
	59		*Harm*		In what ways could data processing result in informational harm on the data subject(s)? (please clarify how such harm could occur, who would be affected and how)	
					How might data processing give rise to any form of	

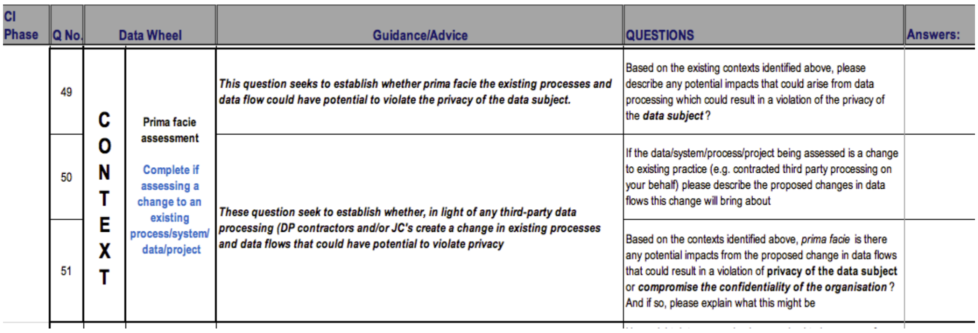
Figure 8.14: Overview - Wider Context

First of all, to address the issue of practitioners potentially 'skipping' the wider context section, this worksheet was amalgamated into the overview section. The way this was done, was to add an additional section at the end of the overview entitled *"Surrounding/Wider Context"*. This way, these questions will be answered as part of the overview and therefore, practitioners are less likely to skip this. In addition a column was added for noting the norm or value to be considered for each question see Figure 8.14 and a section



CI Phase	Q No.	Data Wheel		Guidance/Advice	QUESTIONS
					confidential and not processed? (if yes, please explain)
	66	Positive Impact Assessment	*Data Subject*	***Positive Impact Assessment: note any positive values that could be derived from the data processing***	What is/are the desired effects of the processing for the data subject?
	67				What are the benefits and positive values for the data subject that processing the data will bring to enhance the outcome for the data subject e.g. to facilitating treatment, supply goods, improve mobility, increase transparency etc.
	68		*Organisation*		What is the desired effect of the processing for the organisation?
	69				What are the benefits and positive values that data processing will bring/enhance for the organisation. These may include commercial gain, improved transparency, meeting legal obligation etc.
	70		*Organisation*	***Overriding reason for processing despite risks or lack of consent***	If consent has not been obtained from the data subject, or they have withdrawn their consent, are there any overriding legal, moral or ethical reasons why data processing should be allowed despite this? (e.g. prevention of terrorism, safeguarding of the data subject etc.)
	71				Please note any overriding legal, moral or ethical reasons why data processing should be allowed even if there is a risk of re-identification?

Figure 8.15: Overview - Positive Impact Assessment



CI Phase	Q No.	Data Wheel		Guidance/Advice	QUESTIONS	Answers:
	49	CONTEXT	Prima facie assessment Complete if assessing a change to an existing process/system/ data/project	***This question seeks to establish whether prima facie the existing processes and data flow could have potential to violate the privacy of the data subject.***	Based on the existing contexts identified above, please describe any potential impacts that could arise from data processing which could result in a violation of the privacy of the ***data subject***?	
	50			***These question seek to establish whether, in light of any third-party data processing (DP contractors and/or JC's create a change in existing processes and data flows that could have potential to violate privacy***	If the data/system/process/project being assessed is a change to existing practice (e.g. contracted third party processing on your behalf) please describe the proposed changes in data flows this change will bring about	
	51				Based on the contexts identified above, *prima facie* is there any potential impacts from the proposed change in data flows that could result in a violation of **privacy of the data subject** or ***compromise the confidentiality of the organisation***? And if so, please explain what this might be	

Figure 8.16: Overview - Existing Context

added at the end of the overview to accommodate the Positive Impact Assessment, see Figure 8.15.

Second, because a number of the questions relating to the *prevailing* or *existing* context have been moved to the *"Data Subject"* section, this section now only contains 3 questions, see Figure 8.16. Thus, the final order of the worksheets in the PACT framework will be as depicted in Figure 8.17.

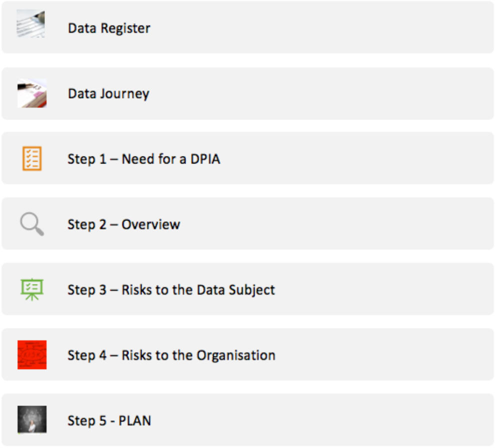


Figure 8.17: PrivACy Throughout (PACT)

8.8 Summary of Findings - PACT

In this Chapter the frameworks created in previous Chapters (6 and 7), were amalgamated to create one complete framework, PrivACy Throughout (PACT), a paper prototype spreadsheet (see Chapter 3, Section 3.4.11), that encompasses a complete privacy risk assessment and planning framework that practitioners can use to support any decision around privacy they need to make.

As part of this, to address the findings from the DPIA Case study (see Chapter 7, Section 7.4) that practitioners fail to differentiate or fully understand the differing perspectives from which privacy risk assessments need to be conducted, the risk assessment element of the DPIA Data Wheel (Chapter 7, Section 7.3.4) was split into two separate risk assessments. In PACT, privacy risks are assessed separately for each perspective: one assesses risk from the data subject's perspective (see Section 8.4.6) and one that assesses risk from the organisation's perspective (see Section 8.4.7). Moreover, a lot more guidance and direction has been provided including a new, more detailed scoring mechanism (see Section 8.4.6 and 8.4.7).

The PACT framework also includes a new section, the privacy lifecycle plan (PLAN), a data governance support paper prototype, devised to help practitioners plan for how they will manage the data during its lifecycle with the organisation. This PLAN has been devised around the 7-stage Data Lifecycle (see Figure 8.7), a data lifecycle model based on the organisational decision model devised in Chapter 2, Section 2.5. It also incorporates a series of ten privacy goals that practitioners need to consider as part of their data management planning (see Table 8.2, Questions 7DS15-27). The PLAN worksheet was evaluated separately along with the revised privacy risk assessments by 9 reviewers from 3 different sectors; academics, practitioners and student peers (see Section 8.5.5, before being incorporated into the main PACT framework.

The whole PACT framework was then evaluated by three different industry sectors, the private sector (see Section 8.6.1); the public sector (see Section 8.6.2); and the charitable sector (see Section 8.6.3). The evaluation resulted in a number of changes being made to improve the framework prior to it being adopted as the privacy assessment framework for the Ideal-Cities EU project (see Figure 8.11).

PACT demonstrates how the privacy-specific frameworks created in earlier work, the meta-model created in Chapter 5, that informed CLIFOD in Chapter 6 (trialled and applied in the *public open sector* domain), and the DPIA Data Wheel in Chapter 7 (applied and trialled in the charity sector, thus confirming the domain neutrality of the framework) can be amalgamated to create one privacy-specific risk assessment framework that practitioners can use to support any privacy decision making the organisation needs to consider, thereby answering RQ3 (see Chapter 1, Section 1.2).

PACT is a prototype spreadsheet for a tool that can be used to assess the privacy impacts and risks of any new or existing dataset, project, system, or process from either the data subject or the organisation's perspective or both. Thus, this prototype spreadsheet confirms proposition P3, a single tool can be created that can facility all three aspects to comprehensive privacy-specific decision making support see Chapter 1, Section 1.2.1). Further, PACT also includes a data lifecycle planning tool (PLAN, see Section 8.5) that practitioners can use to plan how they will manage the data during its lifecycle with the organisation and, as such, is the main contribution of this thesis.

Chapter 9

Discussion & Conclusion

9.1 Introduction

This thesis depicts the journey of creating the PrivACy Throughout (PACT) framework through a series of case studies. Starting with a review of existing theory in Chapter 2, in Chapter 3, the case study was selected as the overarching methodology for this body of work, following Yin (2013) (see Section 3.5).

The main research question this thesis asked was: *”How can privacy assessment be incorporated into organisational decision-making in a practical manner, encompassing legal and contextual considerations, to provide repeatable, effective decision support for determining privacy risks and facilitating the integration of the privacy decision-making function into organisational decision-making by default?”* (see Section 1.2).

To support answering this main research question, three propositions (P’s) and three sub-questions (RQ’s) were asked:

RQ1 *How can existing risk and/or privacy decision frameworks or guidelines be adapted to create a privacy-specific decision-making framework that practitioners can adopt to support privacy decision making?*

RQ2 *How can contemporary legislation be incorporated into the privacy-specific assessment framework (PAF) to practically support practitioners in privacy decision making?*

RQ3 *How can the privacy-specific assessment framework be adapted into a tool that can support any privacy decision making the organisation need to consider?*

P1 *There are existing framework(s) that singularly or through amalgamation of concepts can be adapted to provide a practical foundation for determining privacy risks*;

P2 *The privacy assessment framework (PAF) developed as part of this study incorporates contemporary legislation and enables practitioners to systematically assess privacy risks in practice*;

P3 *A single tool or prototype can be created, based on the PAF, that can facilitate comprehensive privacy-specific decision making support.*

The motivation behind this work was to explore how organisations handle a data subject's privacy when handling their data and, in particular, government organisations (*public bodies*), as they have a legal obligation to publish data in open format (see Section 2.3). The aim of the work was to create a practical solution that will help support practitioners in making consistent and better informed decisions around privacy, thereby providing assurance for end users (*"data subjects"*) that their trust in how an organisation handles their data, or data about them, is not misplaced (see Section 1.1).

The domain chosen to test the framework chosen for adaptation was determined to be the *open data* domain (see Section 1.1.1, in particular *public open data* (see Section 2.3.1), because of the perceived additional risks associated with making data freely available with no restrictions and the fact that public bodies have an obligation to publish data in open format.

The literature review (Chapter 2) established that *organisation man* (see Section 2.5.1), which it was argued, is likely to be our practitioners, makes decisions based on the best information available. To this end a risk based approach is best suited as risk forms an integral part of organisational decision making and that organisations are comfortable adopting a risk based approach to their decision making (see Section 2.6). Therefore, this approach was chosen to provide a basis upon which to build the privacy-specific assessment framework (PAF) created for this work. To support this, the Contextual Integrity (CI) framework devised by Nissenbaum (Nissenbaum 2010) was chosen as the underpinning theory for adaptation into the privacy assessment framework (see Section 2.7.6).

Following an interpretivism paradigm (see Section 3.2.3), the chosen methodology for this work was a case study (Yin 2013) as this method was flexible enough to allow for multiple methods and levels of analysis within a single study (see Section 3.5).

9.2 Summary of findings

To determine a suitable starting point for the best approach to answer the sub-questions, a series of scoping studies were conducted in Chapter 4. These were devised to allow enquiry into the open data domain, in particular, public open data, and how public bodies have reacted to the obligations placed on them to publish data in open format. These studies confirmed that at least one public body sector, local authorities (LAs), do not have standard processes in place for making decisions about the suitability of data for open data publication (see Section 4.6.1). Further, the findings from these initial studies also suggested that no standardised privacy assessment framework existed to assist practitioners in determining what privacy risks might be associated with publishing open data.

To further investigate how privacy risk assessment can be built into organisational decision making, a meta-model of CI, the selected underpinning theory, was created in Chapter 5. The intention was that, by demonstrating how CI could be broken down into key areas based on the theoretical discussion in the book (Nissenbaum 2010), a systematic decision-flow could be established that can be applied in a methodical format for privacy decision support (see Section 5.6). In devising this model, proposition 1 (P1) was confirmed, an existing framework can be adapted to become *”a practical foundation for determining privacy risks”* (P1, see Section 9.1).

From this, CLIFOD (ContextuaL Integrity For Open Data in practice), a more detailed interpretation of CI, was devised in Chapter 6. CLIFOD illustrated how the meta-model (and therefore CI) can be adapted to support strategic privacy decision making for open data publishing in the public open data domain, thereby answering RQ1. CLIFOD took the form of a 98 question prototype spreadsheet. It was evaluated in a real setting as part of a case study working with a UK public body, a LA (see Section 6.5). This study found that the LA participating in the case study did not have any formal privacy assessment framework in place (see Section **??**). Further, all of the datasets assessed (which had already been published as open data) as part of the evaluation of the CLIFOD framework, were found to contain personal and/or sensitive data (see Sections 6.6.1 - 6.6.3). These results therefore confirmed that, although the LA claimed to comply with data protection laws, existing practice was not sufficiently robust to identify privacy risks prior to publication. This finding would indicate that, not only can the meta-model be successfully applied in practice to support privacy decision making, in doing so, the framework also demonstrates how using CLIFOD is likely to result in more informed decisions being reached as a result. Thus, CLIFOD provides a useful privacy decision making framework for the open data publication domain that can, if applied prior to publication, reduce the likelihood of personal data being published inadvertently (see Section 6.7). However, it was clear that this work could equally be applied to other domains where data privacy needs to be considered.

Therefore, the next study, looked at data privacy for any data, not just public open data. This second case study was a GDPR implementation action intervention case study, working with a local Charity in Chapter 7. As part of this case study, three prototype spreadsheets were created; a user story spreadsheet (Section 7.3.1); a master data register (MDR, see Section 7.3.2); and the ’life of the form’ spreadsheet (Section 7.3.2), used for gathering details about the data journey(s) the data will go on during its lifecycle with the organisation.

The main contribution from this case study was a data protection impact assessment (DPIA) framework, *the DPIA Data Wheel*. The DPIA Data Wheel framework was designed specifically with charities in mind (see Section 7.3.4) but is equally applicable for use in any organisation who needs to conduct DPIAs. This DPIA answered RQ2 by incorporating contemporary legislation, GDPR, into the privacy-specific framework being created as part of the overarching work in answering the research questions (see Section 9.1). The final DPIA Data Wheel comprised a legally compliant DPIA process that incorporated two of the prototype spreadsheets (the MDR and the life of the form) from the GDPR implementation, in order to facilitate organisations being able to collate appropriate evidence for demonstrating compliance with GDPR (see Section 2.4.4). Further, the framework questions integrated most of the questions from CLIFOD, including all the contextual questions, into the framework (see Appendix 7.3.4). Finally, the DPIA framework also incorporated a privacy risk assessment, designed to aid practitioners in determining what privacy risks might be associated with data processing associated with the implementation of a new system, project or process or other data processing activity (see Section 7.4).

The DPIA Data Wheel was empirically evaluated: first, with industry and academic peer groups; second, with the participant Charity staff and volunteers through meetings and a series of training sessions; and third, at a workshop with charity industry peers (Section 7.3.4). From these evaluations, 88 privacy risks were elicited and appropriate mitigation strategies identified. However, the findings also highlighted that all of the risks identified from the sessions, were derived from the perspective of the organisation, rather than the individual, despite direction and guidance provided, indicating further work would need to be done to encourage looking at risks from both perspectives. The DPIA Data Wheel framework itself was well received by participants who felt this framework was an effective tool for conducting standardised, repeatable privacy risk assessments that comply with GDPR. Thus, the DPIA Data Wheel confirmed P2 (see Section 9.1) by facilitating legally compliant, systematic privacy risk assessment (Section 7.4).

The final study conducted is described in Chapter 8. This Chapter saw the amalgamation of the two previous frameworks created and added a final element, a Privacy Lifecycle plANning tool, PLAN (see Section 8.5). As part of this, the elements from CLIFOD (Chapter 6) and the DPIA Data Wheel (Chapter 7) were merged into one framework that encompasses privacy decision making in context (Sections 8.4.3 and 8.4.5) for any existing or new project, dataset or system, from both the perspective of the data subject (Section 8.4.6), and the organisation (Section 8.4.7). Moreover, the PACT framework incorporates a forward planning tool, the PLAN (see Section 8.5) that will allow the organisation to create a data management plan that encompasses the whole of the data lifecycle (see Figure 8.7).

The PACT framework has been designed to incorporate the lessons learnt from previous studies to create an overarching privacy risk assessment tool for any organisation who need to determine what risks are associated with either sharing or publishing data and/or a particular practice, system or processing activity and then PLAN how to ensure the data is to be safeguarded during its lifecycle with the organisation (see Figure 8.17). In doing so, this study answered RQ3 by creating the final contribution of this thesis, a privacy risk framework, PrivACy Throughout (PACT), thereby also affirming P3 (see Section 9.1) and demonstrating that a single prototype can be created to support any privacy decision making an organisation needs to make (see Section 9.1).

Thus, this thesis makes four contributions:

1. *Meta-model*, devised a meta-model of contextual integrity where the main elements of the theory have been separated out into a visual representation of how the framework can be applied in a practical setting;
2. *CLIFOD*, a step-by-step prototype spreadsheet for applying CI in practice that practitioners can use to support standardised and informed privacy risk decision making;
3. *The DPIA Data Wheel*, another stand-alone prototype spreadsheet, suitable for conducting consistent, repeatable and informed data protection impact assessments (DPIAs), in line with contemporary legislation, the General Data Protection Regulation (GDPR); and
4. *PACT*, the main contribution of this work, the PACT (PrivACy Throughout) framework, also presented as a prototype spreadsheet.

9.3 Discussion

Organisations go to great lengths to safeguard their assets including their data by, for example, investing heavily in network protection and monitoring systems and installing protective systems such as firewalls and antivirus protection. As part of this, organisations will most likely consider privacy in terms of the security risks associated with a potential data breach, which, if not detected or avoided, can have costly repercussions for the organisation (Martin et al. 2014). Moreover, individuals want assurances that their privacy has been preserved before data is released, rather than after the breach has occurred and this is what motivated this research. How can organisations provide such assurances and ensure that privacy risks have been properly identified and addressed before data is released?

This work argues that privacy should not only be considered as part of security risks, it needs to be considered as part of a separate, privacy-specific risk assessment taking into account all the different aspects of privacy including the data, the actors, the transmission principles and the context and that is what this work has done by creating PACT.

The initial scoping studies conducted showed that public bodies have not, so far, been particularly proactive in publishing raw datasets as open data. Most public bodies did however, comply with FOI and released data under the FOI publication scheme proactively (see Chapter 4, Sections 4.2 and 4.3). With regards to ensuring confidentiality, while one public body provided assurances that they would *”never put out personalised data.” (P1)* (see Section 9.1), the technical team admitted during the interviews that: *”we quality check it [the data] in terms of formats etc., not so much in terms of data being actually correct” (P3)*. This therefore, would indicate that they cannot be sure that privacy has been preserved, a sentiment echoed by another practitioner who claimed that he feared that personal data could have been released already, being *”almost convinced...[if checked that they would find] something we’d missed” (P2)* (see Chapter 4, Section 4.4.2). Further, the CLIFOD trial findings confirmed that this fear was not unfounded when identifiers and quasi-identifiers were found to be present in datasets already published as open data (see Chapter 6, Section 6.6). These findings confirmed that there was a gap in existing processes which, if not addressed, would highly likely result in increased privacy risks, a finding also confirmed by the increasing number of fines and reported data breaches received by the ICO in 2017 (ICO 2017).

This paved the way for this work to aim towards creating a potential solution to this problem. This solution started by answering RQ1 (see Section 9.1) through creating a meta-model of CI which provided proof of concept that an existing privacy framework could be adapted to be used for privacy risk assessment. This was further expanded on in CLIFOD, an exemplar for how the meta-model could be adapted into a practical framework that practitioners can use in practice to support privacy decision making for open data publication.

Then, when the ICO reported, just after GDPR came into force, for the second quarter of 2018, that they had received 4,056 reported data breach incidents (aka cyber incidents) (ICO 2018d), up from 678 incidents in Q4 of 2017 (ICO 2017), this suggested that the increase in reported data breaches was only likely to continue to increase rapidly. These figures represent nearly a three-fold increase on Q4 from the previous year, suggesting that the introduction of compulsory breach notification within 72 hours of a *suspected* data breach brought in by GDPR (see Section 2.4.4), has dramatically increased the number of reported incidents, making organisations more accountable but also more vulnerable if they do not have appropriate processes in place for protecting the privacy of data subjects. This work therefore decided to use the GDPR legislation as the contemporary legislation to incorporate into the privacy framework being created and answer RQ2 and confirm P2 (see Section 9.1), resulting in the DPIA Data Wheel (see Chapter 7).

In creating the DPIA Data Wheel therefore, the privacy-specific framework initially created to assess privacy risk of making public data available in open format (see Chapter 6), now incorporated privacy risk assessment of *any* data, project or system whether it originated from the public, private or charitable sector. Moreover, the DPIA framework is flexible enough to be applied to organisations of all sizes for both assessing privacy risks and for implementing and amassing demonstrable evidence of GDPR compliance (see

Section 7.4).

For the final study, RQ3 and P3 were answered (see Section 9.1) by amalgamating the frameworks already created as part of this work into one overarching framework, PrivACy Throughout (PACT, see Chapter 8). PACT was created was devised to accommodate assessing what privacy risks might be associated with collecting, processing and potentially reusing data from both the perspective of the individuals whose data is being handled (*the data subjects*), and the perspective of the organisations who handle the data. Further, PACT was extended to include a data lifecycle planning tool (PLAN, see Section 8.5), devised to assist practitioners in creating a data management plan for the data from collection through to destruction. Thus, PACT demonstrates how privacy assessment can be embedded into organisational decision making in a practical manner in one prototype tool, PACT, that incorporates both legal and contextual considerations to support standardised, effective decision making, thereby integrating the privacy decision-making function into organisational decision-making by default and answering the main research question (see Section 9.1) and, as such, is the main contribution of this thesis.

PACT has already been adopted as the privacy framework for the Ideal-Cities project (see Figure 8.11). Further, the work had come full circle to achieve circularity, starting with assessing the privacy risks of publishing open data, moving on to assessing the privacy risks of any data, project or system before circulating back to the start to consider the implications of data re(use) in the circular economy through Ideal-Cities (Section 8.8).

9.4 Future Work

Future work will look at how to turn the prototype spreadsheet of PACT into an automated tool to make privacy assessment more accessible and easier to use. For example, different sections that are repeated in the framework multiple times such as the legal basis for data processing (see Section 2.4.4), which in PACT, appears in: the data register (Appendix F, Section F.2.1); the overview (Section 8.4.3); and the PLAN (Section 8.5), can be automated to auto-populate once the data has been entered the first time in one of the sections. Similarly, where a question consists of multiple sections or is layered depending on the answers provided, if a negative answer is entered which means it is not necessary to complete the remainder of the section (e.g. if no DPIA is required), those questions will not be displayed (i.e. the questions relating to risk to the data subject (Step 4, see Section 8.4.6) will not be asked where a DPIA is not required). By automating these and similar elements within the PACT framework, in the opinion of the researcher, the timing, which currently runs at around 5-6 hours to complete all the elements according to the evaluations (see Section 8.6), could be at least halved. It will also increase data validation and automatically highlight issues by flagging them as they occur.

As part of this work, an investigation will also be conducted to ascertain how the CAIRIS security threat modelling framework (Faily et al. 2012) and the PIA process devised by Alshammari and Simpson (2018) can also be incorporated into such tool support to help elicit and model privacy threats and identify vulnerabilities.

Bibliography

Abiteboul, S., Hull, R. and Vianu, V., eds., 1995. *Foundations of Databases: The Logical Level*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1st edition.

Acito, F. and Khatri, V., 2014. Business analytics: Why now and what next? *Business Horizons*, 57 (5), 565 – 570.

Ackoff, R., 1989. From data to wisdom. *Journal of Applied Systems Analysis*, 16 (3), 3–9.

Alshammari, M. and Simpson, A., 2018. Towards an effective pia-based risk analysis: An approach for analysing potential privacy risks. [online].

Altman, M., Wood, A., O'Brien, D. R., Vadhan, S. and Gasser, U., 2015. Towards a modern approach to privacy-aware government data releases. *Berkeley Technology Law Journal*, 30 (3), 1967.

ANSI, 2011. ANSI/ASSE Z690.1-2011 Vocabulary for Risk Management (National Adoption of: ISO Guide 73:2009). Technical report, American Society of Safety Engineers ANSI/ASSE.

Arghode, V., 2012. Qualitative and quantitative research: Paradigmatic differences. *Global Education Journal*, 2012 (4), 155–163.

Arksey, H. and O Malley, L., 2005. Scoping studies: Towards a methodological framework. *INTERNATIONAL JOURNAL OF SOCIAL RESEARCH METHODOLOGY*, 8 (1), 19 – 32.

Arthur, C. and Cross, M., 2015. Free our data: Make taxpayers' data available to them. [online]. URL `http://www.freeourdata.org.uk/`.

Bamberger, K. A. and Mulligan, D. K., 2015. *Privacy on the Ground: Driving Corporate Behaviour in the United States and Europe*. London: England: the MIT Press: Mssachusetts Institute of Technology.

Barth, A., Anupam, D., Mitchell, J. C. and Nissenbaum, H. F., 2006. Privacy and contextual integrity: Framework and applications. *Proceedings - IEEE Symposium on Security and Privacy*, volume 2006, 184–198.

Bartling, J., 2015. The enron data set - where did it come from? *Bartling Forensic*, [online].

Bauer, L., Cranor, L. F., Reeder, R. W., Reiter, M. K. and Vaniea, K., 2009. Real life challenges in access-control management. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, CHI '09, 899–908.

Baxter Magolda, M. B., 2004. Evolution of a constructivist conceptualization of epistemological reflection. *Educational Psychologist*, 39 (1), 31–42.

Bernard, H. R., 1988. *Research methods in cultural anthropology*. Newbury Park, Calif.: Sage Publications.

Berners-Lee, T., 2006. 5* open data. [online]. URL `https://www.w3.org/DesignIssues/LinkedData.html`.

Bettini, C. and Riboni, D., 2015. Privacy protection in pervasive systems: State of the art and technical challenges. *Pervasive and Mobile Computing*, 17 (Part B), 159 – 174.

Beyer, H. and Holtzblatt, K., 1998. *Contextual design : defining customer-centered systems.*. San Francisco, Calif. : Morgan Kaufmann Publishers, c1998.

Bhushan, N. and Rai, K., 2007. *Strategic decision making: applying the analytic hierarchy process*. London: England: Springer Science & Business Media.

Bissonette, M., 2016. *Project Risk Management : A Practical Implementation Approach.*. Newtown Square, Pennsylvania: Project Management Institute.

Blackburn, M., 2016. A day for consumer awareness. *The Hindu Business Line*, [online].

Blaxter, L., Hughes, C. and Tight, M., 2006. *How to research.*. [electronic resource]., Berkshire, England ; New York :: Open University Press, 3rd edition.

Blaxter, L., Hughes, C. and Tight, M., 2010. *How to research. [electronic resource].*. Open UP study skills, Maidenhead : Open University Press, 2010.

Borgesius, F. Z., Gray, J. and Eechoud, M. V., 2015. Open data, privacy, and fair information principles: Towards a balancing framework. *Berkeley Technology Law Journal*, 30 (3), 2073.

Borghi, M., Ferretti, F. and Karapapa, S., 2013. Online data processing consent under eu law: A theoretical framework and empirical evidence from the uk [article]. *International Journal of Law and Information Technology*, 21 (2), 109.

Boswarva, O., 2016. Uk local government open data resources (may 2016). [online]. URL `https://www.owenboswarva.com/localopendata.htm`.

Boyd, D. and Crawford, K., 2012. Critical questions for big data. *Information, Communication & Society*, 15 (5), 662 – 679.

Boylan, M., 2009. *Critical Inquiry : The Process of Argument.*. [ebook], Westview Press.

Bruner, J. S., 1986. *Actual minds, possible worlds. [electronic resource].*. Cambridge, MA : Harvard University Press, 1986.

BS ISO 27000:2017, 2017. British standards document bs iso 27000:2017: Information technology. security techniques. information security management systems. overview and vocabulary. Technical report, British Standard and the International Organization for Standardization (ISO).

BS ISO 31000, 2009. British standards document bs iso 31000:2009: Risk management. principles and guidelines. Technical report, British Standard and the International Organization for Standardization (ISO).

BS ISO/IEC 29100, 2011. BS ISO/IEC29100: Information technology — security techniques — privacy framework. Technical report, British Standard and the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC).

Burton, J., 2002. Mind the gap. *The New Law Journal*, 152 (152 NLJ 1933).

Cabinet Office, 2015a. Open standards principles. [online]. URL `https://www.gov.uk/government/publications/open-standards-principles/open-standards-principles#principles-for-selecting-open-standards`.

Cabinet Office, 2015b. The National Information Infrastructure (NII) Implementation Document. [online]. URL `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/416472/National_Infrastructure_Implementation.pdf`.

Cagliano, A. C., Grimaldi, S. and Rafele, C., 2015. Choosing project risk management techniques. a theoretical framework. *Journal of Risk Research*, 18 (2), 232 – 248.

Care Quality Commission (CQC), 2018. Care quality commission. [online]. URL `https://www.cqc.org.uk/`.

Carrara, W., Chan, W. S., Fisher, S. and van Steenbergen, E., 2015. Creating value through open data: Study on the impact of re-use of public data resources. Technical report, Capgemini Consulting on behalf of the European Commission (EC): European Data Portal.

Cavoukian, A., 2011. Privacy by design: The 7 foundational principles. Technical report, Information and Privacy Commissioner of Ontario, Ontario: Canada.

Chen, P. P.-S., 1976. The entity-relationship model toward a unified view of data. *ACM Trans. Database Syst.*, 1 (1), 9–36.

Christie, M., Carey, M., Robertson, A. and Grainger, P., 2015. Putting transformative learning theory into practice. *Australian Journal of Adult Learning*, 55 (1), 9–30.

Chui, M., Farrell, D. and Jackson, K., 2014. How government can promote open data. Technical report, McKinsey and Company. URL https://www.mckinsey.com/industries/public-sector/our-insights/how-government-can-promote-open-data.

CNIL, 2012. Methodology for risk management. Technical report, Commission Nationale de l'Informatique et des Libertés (CNIL), Paris: France. URL https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf.

Cohen, D. and Crabtree, B., 2006. Qualitative research guidelines project. [online].

Conley, A., Datta, A., Helen, N. and Sharma, D., 2012. Sustaining privacy and open justice in the transition to online court records: A multidisciplinary inquiry. *Maryland Law Review*, 71 (3), 772 – 847.

Corbin, J. M. and Strauss, A. L., 2008. *Basics of qualitative research : techniques and procedures for developing grounded theory.*. Los Angeles, Calif. ; London : SAGE, c2008.

CORDIS, 2017. Intelligence-driven urban internet-of-things ecosystems for circular, safe and inclusive smart cities (ideal-cities). [online]. URL https://cordis.europa.eu/project/rcn/213015/factsheet/en.

Council of Europe, 1950. European convention on human rights, as amended by protocols nos. 11 and 14. Technical report, Council of Europe.

Cowling, K. and Sugden, R., 1998. The essence of the modern corporation: Markets, strategic decision-making and the theory of the firm. *Manchester School (14636786)*, 66 (1), 59.

Creswell, J. W., 1998. *Qualitative inquiry and research design. [electronic resource] : choosing among five traditions.*. Thousand Oaks, Calif. ; London : Sage, 1998.

Creswell, J. W., 2009. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. Thousand Oaks : Calif: Sage Publications, 3rd edition.

Daley, J., 2016. Driven by data. *Entrepreneur*, 44 (1), 133 – 139.

Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Métayer, D. L., Tirtea, R. and Schiffner, S., 2014. Privacy and data protection by design privacy and data protection by design – from policy to engineering. Technical report, European Union Agency for Network and Information Security (ENISA).

Dardenne, A., van Lamsweerde, A. and Fickas, S., 1993. Goal-directed requirements acquisition. *Science of Computer Programming*, 20 (1), 3 – 50.

David, W., Rachel, F. and Rowena, R., 2013. A comparative analysis of privacy impact assessment in six countries. *Journal of Contemporary European Research, Vol 9, Iss 1, Pp 160-180 (2013)*, 9 (1), 160 – 180.

Davison, G. C., Vogel, R. S. and Coffman, S. G., 1997. Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. *Journal Of Consulting And Clinical Psychology*, 65 (6), 950 – 958.

Demirci, A. E., 2016. Change-specific cynicism as a determinant of employee resistance to change. *Is, Guc: The Journal of Industrial Relations and Human Resources*, 18 (4), 1 – 20.

Deng, M., Wuyts, K., Scandariato, R., Preneel, B. and Joosen, W., 2011. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16 (1), 3–32.

Department for Communities and Local Government, 2015. Local government transparency code 2015. [online]. URL `https://www.gov.uk/government/publications/local-government-transparency-code-2015`.

Desilets, A., 2008. Tell me a story. *IEEE Software*, 25 (2), 14–15.

Duhigg, C., 2005. The odds of disaster. [online].

Dwork, C., 2006. Differential privacy. *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, Berlin, Heidelberg: Springer Berlin Heidelberg, volume 4052 of *Lecture Notes in Computer Science*, 1–12.

Edwards, W., 1954. The theory of decision making. *Psychological bulletin*, 51 (4), 380–417.

EEA, 2016. Circular economy in europe: Developing the knowledge base. [online]. URL `https://www.eea.europa.eu/publications/circular-economy-in-europe`.

EEA, 2017. Circular by design: Products in the circular economy. [online]. URL `https://www.eea.europa.eu/publications/circular-by-design`.

Eisenhardt, K. M., 1989. Building theories from case study research. *Academy of Management Review*, 14 (4), 532 – 550.

El Emam, K., Jonker, E., Arbuckle, L. and Malin, B., 2011. A systematic review of reidentification attacks on health data. *Plos One*, 6 (12), e28071.

Ericsson, K. A. and Simon, H. A., 1980. Verbal reports as data. *Psychological Review*, 87 (3), 215–250.

Europa, 2018. EU Law: Decision Making: Regulations, Directives and other acts. [online]. URL `https://europa.eu/european-union/eu-law/legal-acts_en`.

European Parliament and the Council of Europe, 2016. General data protection regulation (gdpr). Regulation (EU) 2016/679 5419/1/16, European Parliament and the Council of Europe, Brussels.

Faily, S., 2018. *Designing Usable and Secure Software with IRIS and CAIRIS*. Springer. In Press.

Faily, S. and Fléchais, I., 2009. Context-Sensitive Requirements and Risk Management with IRIS. *Proceedings of the 17th IEEE International Requirements Engineering Conference*, IEEE Computer Society, 379–380.

Faily, S. and Fléchais, I., 2010a. A Meta-Model for Usable Secure Requirements Engineering. *Proceedings of the 6th International Workshop on Software Engineering for Secure Systems*, IEEE, 126–135.

Faily, S. and Fléchais, I., 2010b. Analysing and Visualising Security and Usability in IRIS. *Proceedings of the 5th International Conference on Availability, Reliability and Security*, IEEE, 543–548.

Faily, S. and Fléchais, I., 2010. Barry is not the weakest link: Eliciting secure system requirements with personas. *Proceedings of the 24th BCS Interaction Specialist Group Conference*, Swinton, UK, UK: British Computer Society, BCS '10, 124–132.

Faily, S. and Fléchais, I., 2010. Towards tool-support for Usable Secure Requirements Engineering with CAIRIS. *International Journal of Secure Software Engineering*, 1 (3), 56–70.

Faily, S., Lyle, J., Fléchais, I. and Simpson, A., 2015. Usability and Security by Design: A Case Study in Research and Development. *Proceedings of the NDSS Workshop on Usable Security*, Heriot Watt University Edinburgh, Internet Society.

Faily, S., Lyle, J., Namiluko, C., Atzeni, A. and Cameroni, C., 2012. Model-driven architectural risk analysis using architectural and contextualised attack patterns. *Proceedings of the Workshop on Model-Driven Security*, ACM, 3:1–3:6.

Farrell, S., 2015. Nearly 157,000 had data breached in talktalk cyber-attack. [online].

FERMA, 2003. A risk management standard. Technical report, Federation of European Risk Management Associations (FERMA), Belgium: Brussels.

Financial Conduct Authority, 2018. Fca and ico publish joint update on gdpr. [online]. URL https://www.fca.org.uk/news/statements/fca-and-ico-publish-joint-update-gdpr.

Fiordelisi, F., Soana, M.-G. and Schwizer, P., 2013. The determinants of reputational risk in the banking sector. *Journal of Banking and Finance*, 37, 1359 – 1371.

Fishenden, J. and Thompson, M., 2013. Digital government, open architecture, and innovation: Why public sector it will never be the same again. *Journal of Public Administration Research & Theory*, 23 (4), 977.

Fleming, J. E., 1966. Study of a business decision. *California Management Review*, 9 (2), 51 – 56.

Fonteyn, M., Kuipers, B. and Grobe, S., 1993. A description of think aloud method and protocol analysis. *Qualitative Health Research*, 3 (4), 430 – 441.

Fowler, M., 2004. *UML distilled : a brief guide to the standard object modeling language.*. The Addison-Wesley object technology series, Boston, MA. : Addison-Wesley.

Fried, C., 1970. *An anatomy of values; problems of personal and social choice*. Harvard University Press.

Friedmann, W., 1951. The legal status and organization of the public corporation. *Law and Contemporary Problems*, 16 (4), 576 – 593.

Fung, B. C. M., Ke, W., Rui, C. and Yu, P. S., 2010. Privacy preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42 (4), 14:1–14:53.

G8, 2013. G8 Open Data Charter. [online]. URL https://opendatacharter.net/g8-open-data-charter/.

Gagnon, Y.-C., 2010. *The Case Study As Research Method : A Practical Handbook.*. Les Presses de l'Université du Québec.

Galanc, T., Kolwzan, W., Pieronek, J. and Skowronek-Gradziel, A., 2016. Logic and risk as qualitative and quantitative dimensions of decision-making process. *Operations Research & Decisions*, 26 (3), 21.

Gavison, R., 1980. Privacy and the limits of the law. *Yale Law Journal*, 89 (3), 421–471.

Geiger, C. P. and von Lucke, J., 2012. Open government and (linked) (open) (government) (data). *eJournal of eDemocracy & Open Government*, 4 (2), 265.

Given, L., ed., 2008. *The SAGE Encyclopedia of Qualitative Research Methods [online]*. Thousand Oaks, California: Sage Publications Inc.

Goda, S., 2011. Open data and open government. *INFORMACIOS TARSADALOM*, 11 (1-4), 181 – +.

Google, 2015. Google drive: Fusion tables. [online]. URL `https://sites.google.com/site/fusiontablestalks/home`.

Gov.UK, 2016a. APIs: Using and creating Application Programming Interfaces. [online]. URL `https://www.gov.uk/service-manual/technology/application-programming-interfaces-apis`.

Gov.UK, 2016b. Blog: Charity commission: Accountability. [online]. URL `https://charitycommission.blog.gov.uk/tag/accountability/`.

Grodzinsky, F. S. and Tavani, H. T., 2011. Privacy in ”the cloud”: Applying nissenbaum's theory of contextual integrity. *SIGCAS Comput. Soc.*, 41 (1), 38–47.

Guba, E. C., 1990. The alternative paradigm dialog. *The paradigm dialog*, Newbury Park, CA: Sage, 17–30.

Gutek, G. L., 2009. *New Perspectives on Philosophy and Education*. Pearson Custom Education Series, Pearson.

Hall, D. C., 2011. Making risk assessments more comparable and repeatable. *Systems Engineering*, 14 (2), 173 – 179.

Hansen, M., 2012. *Top 10 mistakes in system design from a privacy perspective and privacy protection goals*, Springer Berlin Heidelberg, volume 375 of *IFIP AICT*. 14–31.

Hansen, M., Jensen, M. and Rost, M., 2015. Protection goals for privacy engineering. *2015 IEEE Security and Privacy Workshops*, Unabhañgiges Landeszentrum fuř Datenschutz Schleswig-Holstein (ULD), 159–166.

Harrison, E. F. and Pelletier, M. A., 1998. Foundations of strategic decision effectiveness. *Management Decision*, 36 (3), 147–159.

Harrison, E. F. and Pelletier, M. A., 2000. The essence of management decision. *Management Decision*, 38 (7), 462–470.

Heiser, J., 2008. A simple method for expressing information criticality and classification. [online], Gartner.

Henriksen-Bulmer, J., 2016. A Framework for Public Bodies for Managing the Secure and Appropriate Release of Open Source Data. *British HCI 2016 Doctoral Consortium*, Bournemouth University.

Henriksen-Bulmer, J. and Faily, S., 2017. Applying contextual integrity to open data publishing. *Proceedings of the 31st British HCI Group Annual Conference on People and Computers: Digital Make Believe*, British Computer Society.

Henriksen-Bulmer, J., Faily, S. and Jeary, S., 2018. DPIAs for Charities: a Charity Sector Specific DPIA Framework. *IFIP Advances in Information and Communication Technology*, 13th IFIP WG 9.2, 9.6/11.7, 11.6 International Summer School, Vienna, Austria 20-25 Aug 2018. URL `https://www.researchgate.net/publication/327437377_DPIAs_for_Charities_The_DPIA_Data_Wheel`.

Henriksen-Bulmer, J., Faily, S. and Jeary, S., 2019a. *Privacy and Identity Management. Fairness, Accountability, and Transparency in the Age of Big Data*, Switzerland: Springer International Publishing, chapter Implementing GDPR in the Charity Sector: A Case Study. IFIP Advances in Information and Communication Technology, 13th IFIP WG 9.2, 9.6/11.7, 11.6 International Summer School, Vienna, Austria 20-24 Aug 2018, Revised Selected Papers edition, 173–188.

Henriksen-Bulmer, J., Faily, S. and Jeary, S., 2019b. Privacy risk assessment in context: A meta-model based on contextual integrity. *Computers & Security*, 82, 270 – 283. URL `{https://www.sciencedirect.com/science/article/pii/S0167404818301998?via%3Dihub}`.

Henriksen-Bulmer, J., Faily, S. and Katos, V., 2018. Translating Contextual Integrity into Practice using CLIFOD. *Proceedings of the 2018 Networked Privacy Workshop at CSCW*.

Henriksen-Bulmer, J. and Jeary, S., 2016. Re-identification attacks: A systematic literature review. *International Journal of Information Management*, 36, 1184–1192.

Herzberg, B., 2014. The next frontier for open data: An open private sector. *The World Bank*, [online].

Hirsch, D. D., 2013. The glass house effect: Big data, the new oil, and the power of analogy. *Maine Law Review*, 66 (2), 373–396.

Holtzblatt, K., Wendell, J. B. and Wood, S., 2005. *Rapid contextual design. [electronic resource] : a how-to guide to key techniques for user-centered design.*. The Morgan Kaufmann series in interactive technologies, San Francisco : Elsevier/Morgan Kaufmann.

Houghton, C., Hunter, A. and Meskell, P., 2012. Linking aims, paradigm and method in nursing research. *Nurse Researcher*, 20 (2), 34 – 39.

Howard, M. and Lipner, S., 2006. *The Secuirty Development Lifecycle*. 9780735622742, Redmond, WA: Microsoft Press.

Hyland, K., 2004. *Disciplinary discourses: social interations in academic writing*. Michigan: US: The University of Michigan Press, michigan classics edition edition.

ICO, 2012. Anonymisation: managing data protection risk code of practice. [online]. URL `https://ico.org.uk/media/1061/anonymisation-code.pdf`.

ICO, 2017. Ico data security trends: What action we've taken in q4, what you've reported to us and what you can do to stay secure. Technical report, Information Commissioners Office (ICO), London : UK.

ICO, 2017. Preparing for the general data protection regulation (GDPR): 12 steps to take now. Technical Report V2.0 20170525, Information Commissioner's Office. URL `https://ico.org.uk/media/2014146/gdpr-12-steps-infographic-201705.pdf`.

ICO, 2018a. Data protection impact assessments (DPIAs). [online]. URL `https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/`.

ICO, 2018b. For organisations: Guide to the general data protection regulation (gdpr). [online]. URL `https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/`.

ICO, 2018c. General data protection regulation (GDPR) faqs for charities. [online]. URL `https://ico.org.uk/for-organisations/charity/`.

ICO, 2018d. Ico data security trends: What action we've taken in q2, what you've reported to us and what you can do to stay secure. [online]. URL `https://ico.org.uk/action-weve-taken/data-security-incident-trends/`.

Information Commissioners Office, 2013. Freedom of information act definition document for principal local authorities (county councils, unitary authorities, metropolitan district councils borough councils, city councils and district councils, the council of the isles of scilly and local authorities in wales). [online]. URL `https://ico.org.uk/media/1262/definition_document_local_authorities.pdf`.

Information Commissioners Office, 2014. Conducting privacy impact assessments: code of practice. Technical Report 20140225, Information Commissioners Office (ICO). URL `https://ico.org.uk/global/page-not-found?aspxerrorpath=/media/1595/pia-code-of-practice.pdf`.

Information Commissioners Office, 2015a. Datasets (sections 11, 19 & 45): Freedom of information act. [online]. URL `https://ico.org.uk/media/1151/datasets-foi-guidance.pdf`.

Information Commissioners Office, 2015b. Model publication scheme, version 1.2. [online]. URL https://ico.org.uk/media/for-organisations/documents/1153/model-publication-scheme.pdf.

Information Commissioners Office, 2016a. Choosing appropriate formats: Help your users by providing content in a format they can use. [online].

Information Commissioners Office, 2016b. The guide to the re-use of public sector information regulations 2015. [online]. URL https://ico.org.uk/for-organisations/guide-to-rpsi/.

Kallio, H., Pietilä, A.-M., Johnson, M. and Kangasniemi, M., 2016. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, 72 (12), 2954 – 2965.

Kalloniatis, C., Kavakli, E. and Gritzalis, S., 2008. Addressing privacy requirements in system design: the pris method. *Requirements Engineering*, 13 (3), 241 – 255.

Kavakli, E., Kalloniatis, C., Loucopoulos, P. and Gritzalis, S., 2006. Incorporating privacy requirements into the system design process: The pris conceptual framework. *Internet Research*, 16 (2), 140–158.

Kimble, C. and Milolidakis, G., 2015. Big data and business intelligence: Debunking the myths. *Global Business and Organizational Excellence*, 35 (1), 23 – 34.

Krupa, Y. and Vercouter, L., 2012. Handling privacy as contextual integrity in decentralized virtual communities: The privacias framework. *Web Intelligence & Agent Systems*, 10 (1), 105 – 116.

Kulk, B., Stefan; van Loenen, 2012. Brave new open data world? *International Journal of Spatial Data Infrastructures Research*, 7, 196–206.

Lablans, M., Borg, A. and Ückert, F., 2015. A restful interface to pseudonymization services in modern web applications. *BMC Medical Informatics & Decision Making*, 15 (1), 1 – 10.

Lamsweerde, A. V., Brohez, S., Landtsheer, R. D., Janssens, D. and Informatique, D. D., 2003. From system goals to intruder anti-goals: Attack generation and resolution for security requirements engineering. *In Proc. of RHAS'03*, 49–56.

Lederer, S., Mankoff, J. and Dey, A., 2003. Towards a deconstruction of the privacy space. Workshop on Privacy In Ubicomp 2003: Ubicomp communities: privacy as boundary negotiation, 1–7.

Liew, A., 2007. Understanding data, information, knowledge and their inter-relationships. *Journal of Knowledge Management Practice*, 7 (2).

Liew, A., 2013. Dikiw: Data, information, knowledge, intelligence, wisdom and their interrelationships. *Business Management Dynamics*, 2 (10), 49 – 62.

Lipkus, I. and Hollands, J., 1999. The visual communication of risk. *JNCI: Journal of the National Cancer Institute*, 91 (1), 149 – 163.

Lyon, B. K. and Popov, G., 2016a. The art of assessing risk. (cover story). *Professional Safety*, 61 (3), 40 – 51.

Lyon, B. K. and Popov, G., 2016b. The art of assessing risk. (cover story). *Professional Safety*, 61 (3), 40 – 51.

Machara, S., Chabridon, S. and Taconet, C., 2013. Trust-based context contract models for the internet of things. *Ubiquitous Intelligence and Computing, 2013 IEEE 10th International Conference on and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC)*, 557–562.

Martin, C., Kadry, A. and Abu-Shady, G., 2014. Quantifying the financial impact of it security breaches on business processes. *2014 Twelfth Annual International Conference on Privacy, Security and Trust, Privacy, Security and Trust (PST)*, 149.

McAfee, A. and Brynjolfsson, E., 2012. Big data: the management revolution. *Harvard business review*, 90 (10), 60–66, 68, 128.

McDaniels, O. B., 1983. Existentialism and pragmatism: the effect of philosophy on methodology of teaching. *Journal of Nursing Education*, 22 (2), 62 – 66.

Mezirow, J., 1997. Transformative learning: Theory to practice. *New Directions for Adult & Continuing Education*, (Issue 74), 5–12.

Michiel, R., 2015. Big data and consumer participation in privacy contracts: Deciding who decides on privacy. *Utrecht Journal of International and European Law*, 31 (80), 51–71.

Millar, A., Simeone, R. S. and Carnevale, J. T., 2001. Logic models: a systems tool for performance management. *Evaluation and Program Planning*, 24 (1), 73 – 81.

Ministry of Justice, 2013. Secretary of State's Code of Practice (datasets) on the discharge of public authorities' functions under Part 1 of the Freedom of Information Act. URL `https://www.gov.uk/government/publications/secretary-of-states-code-of-practice-datasets-on-the-discharge-of-public-authorities-func`

Mitchell, V.-W. and Nygaard, A., 1999. Consumer perceived risk: Conceptualisations and models. *European Journal of Marketing*, 33 (1-2), 163 – 195.

Moon, J., 2004. Using reflective learning to improve the impact of short courses and workshops. *JOURNAL OF CONTINUING EDUCATION IN THE HEALTH PROFESSIONS*, 24 (1), 4–11.

Moore, A. D., 2015. *Privacy, Security and Accountability : Ethics, Law and Policy.*. Rowman & Littlefield International.

Mulligan, D. K., Koopman, C. and Doty, N., 2016. Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. *Philosophical Transactions. Series A, Mathematical, Physical, And Engineering Sciences*, 374 (2083).

Narayanan, A. and Shmatikov, V., 2008. Robust de-anonymization of large sparse datasets. *2008 IEEE Symposium on Security and Privacy (sp 2008), Security and Privacy, 2008. SP 2008. IEEE Symposium on*, 111.

National Drug Evidence Centre, 2018. National Drug Treatment Monitoring System (NDTMS). [online]. URL `https://www.ndtms.net/`.

National Treasury, 2010. Risk management framework. Technical report, Republic of South Africa.

Nissenbaum, H., 2004. Privacy as contextual integrity. *Washington Law Review*, 79 (1), 119–158.

Nissenbaum, H. F., 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life.*. Stanford: California: Stanford Law Books.

NIST, 2010. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). Technical Report 800-122, National Institute of Standards and Technology (NIST); U.S. Department of Commerce.

NIST, 2012. Guide for conducting risk assessments. Technical Report SP 800-30, National Institute of Standards and Technology (NIST); U.S. Department of Commerce, US: Gaithersburg: MD.

Obama, B., 2009. Transparency and open government: Memorandum for the heads of executive departments and agencies. URL `https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2009/m09-12.pdf`.

OECD, 2013. The OECD Privacy Framework. Technical Report C(80)58/FINAL, Organisation for Economic Co-operation and Development (OECD). URL `https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf`.

Oetzel, M. C. and Spiekermann, S., 2014. A systematic methodology for privacy impact assessments: a design science approach. *European Journal of Information Systems*, 23 (2), 126.

Office of Management and Budget, 2013. Memorandum for the heads of executive departments and agencies: Open data policy - managing information as an asset. URL https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.

Ohm, P., 2010. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review*, 57 (6), 1701 – 1777.

O'Leary, D. E., 1988. Expert system prototyping as a research tool. *Applied Expert Systems*, 17–31.

Open Data Charter, 2015. International open data charter. [online]. URL https://opendatacharter.net/.

Open Data Institute, 2015. Open data means business: Uk innovation across sectors and regions. [online]. URL https://theodi.org/article/open-data-means-business/.

Open Data Institute, 2016. What is open data? [online]. URL http://opendatahandbook.org/guide/en/what-is-open-data/.

Open Government Partnership, 2016. Participating countries. [online]. URL https://www.opengovpartnership.org/participants.

Open Government Working Group, 2007. Open government data principles. [online]. URL https://opengovdata.org/.

Open Knowledge, 2016. Open data index: Tracking the state of government open data. [online]. URL http://2015.index.okfn.org/about/.

Oxford Dictionaries, 2016. English : Oxford living dictionaries. [online]. URL https://en.oxforddictionaries.com/.

Oxford University Press, 2017. Oxford english dictionary. [online]. URL http://www.oed.com/.

Palen, L. and Dourish, P., 2003. Unpacking "privacy" for a networked world. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, CHI '03, 129–136. URL http://doi.acm.org/10.1145/642611.642635.

Parker, R. B., 1974. A definition of privacy. *Rutgers Law Review*, 27 (2), 275–297.

Pfitzmann, A. and Hansen, M., 2010. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. [online]. V0.34.

Pinsonneault, A. and Kraemer, K. L., 1993. Survey research methodology in management information systems: An assessment. *Journal of Management Information Systems*, 10 (2), 75–105.

Piper, F. and Murphy, S., 2002. *Cryptography: A Very Short Introduction*. Very Short Introductions, OUP Oxford.

Po-Ching, L. and Pei-Ying, L., 2016. Unintentional and involuntary personal information leakage on facebook from user interactions. *KSII Transactions on Internet & Information Systems*, 10 (7), 3301.

Ponemon Institute, 2015. 2015 cost of data breach study: Global analysis. [online]. URL `https://securityintelligence.com/cost-of-a-data-breach-2015/`.

Popper, K. R., 2005. *The Logic of Scientific Discovery*. [electronic resource], New York, NY, USA: Routledge Classics.

Project Management Institute, 2004. *A guide to the project management body of knowledge : PMBOK guide.*. Newton Square, Pa. : Project Management Institute, c2004.

Province of British Columbia, 2012. Risk management guideline for the bc public sector. Technical report, Province of British Columbia: Risk Management Branch and Government Security Office.

Public Sector Transparency Board, 2012. Public sector transparency board: Public data principles. [online]. URL `PublicDataPrinciples_ForData.Gov(1)_10.pdf`.

Rooney, T., Lawlor, K. and Rohan, E., 2016. Telling tales: Storytelling as a methodological approach in research. *Electronic Journal of Business Research Methods*, 14 (2), 147 – 156.

Rowley, J., 2002. Using case studies in research. *Management Research News*, 25 (1), 16–27.

Rumbaugh, J., Jacobson, I. and Booch, G., 2004. *Unified Modeling Language Reference Manual, The (2nd Edition)*. Pearson Higher Education.

de Ruyter, K. and Scholl, N., 1998. Positioning qualitative market research: reflections from theory and practice. *Qualitative Market Research: An International Journal*, 1 (1), 7–14.

Samarati, P., 2001. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge & Data Engineering*, 13 (6), 1010.

Sanchez, R. and Heene, A., 2004. *The new strategic management: organization, competition and competence*. New York, NY, USA: Wiley.

Sar, R. K. and Al-Saggaf, Y., 2013. Contextual integrity's decision heuristic and the tracking by social network sites. *ETHICS AND INFORMATION TECHNOLOGY*, 16 (1), 15 – 26.

Sasse, M. A., Brostoff, S. and Weirich, D., 2001. Transforming the 'weakest link' a human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19 (3), 122.

Sawyer, S., 2001. Qualitative research in is. E. M. Trauth, ed., *Qualitative Research in IS: Issues and Trends*, Hershey, PA, USA: IGI Global, chapter Analysis by Long Walk: Some Approaches to the Synthesis of Multiple Sources of Evidence, 163–191.

Schwandt, T. A., 2001. *Dictionary of qualitative inquiry.*. London : UK: Sage, 2nd edition.

Sefelin, R., Tscheligi, M. and Giller, V., 2003. Paper prototyping - what is it good for?: A comparison of paper- and computer-based low-fidelity prototyping. *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA: ACM, CHI EA '03, 778–779.

Shakespeare, S., 2013. Shakespeare review: An independent review of public sector information. URL `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/198752/13-744-shakespeare-review-of-public-sector-information.pdf`.

Shane, D., 2015. Battling for the rights to privacy and data protection in the irish courts. *Utrecht Journal of International and European Law*, 31 (80), 131–136.

Simon, H. A., 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69 (1), 99–118.

Simon, H. A., 1979. Rational decision making in business organizations. *The American Economic Review*, 69 (4), 493–513.

Simpson, A. C., 2011. On privacy and public data: A study of data.gov.uk. *Journal of Privacy &amp; Confidentiality*, 3 (1), 51–65.

Solove, D. J., 2006. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154 (3), 477 – 564.

Spiekermann, S. and Cranor, L., 2009. Engineering privacy. *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*, 35 (1), 67–82.

Sunlight Foundation, 2010. Ten principles for opening up government information. URL `https://sunlightfoundation.com/policy/documents/ten-open-data-principles/`.

Swire, P. P. and Ahmad, K., 2012. *Foundations of Information Privacy and Data Protection: A Survey of Global Concepts, Laws and Practices*. US: Portsmouth, NH: The International Association of Privacy Professionals (IAPP).

The General Assembly of the United Nations, 1948. The Universal Declaration of Human Rights. [online].

The National Archives, 2015. Guidance on the implementation of the reuse of public sector information regulations 2015. [online]. URL `http://www.nationalarchives.gov.uk/documents/information-management/psi-implementation-guidance-public-sector-bodies.pdf`.

Thomson, D., Bzdel, L., Golden-Biddle, K., Reay, T. and Estabrooks, C. A., 2005. Central questions of anonymization: A case study of secondary use of qualitative data. *Forum: Qualitative Social Research*, 6 (1), 1 – 16.

UK Parliament, 2018. Data protection act 2018. [online]. URL `https://services.parliament.uk/bills/2017-19/dataprotection.html`.

UrbanTide, 2016. Open data - is the open private sector the next frontier? [online]. URL `https://urbantide.com/fullstory2/2016/10/24/open-data-is-the-open-private-sector-the-next-frontier`.

Urquhart, C., Lehmann, H. and Myers, M. D., 2010. Putting the ‘theory’ back into grounded theory: guidelines for grounded theory studies in information systems. *Information Systems Journal*, 20 (4), 357 – 381.

Virlics, A., 2013. Investment decision making and risk. *Procedia Economics and Finance*, 6, 169 – 177.

Walker, M., Takayama, L. and Landay, J. A., 2002. High-fidelity or low-fidelity, paper or computer? choosing attributes when testing web prototypes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46 (5), 661–665.

Weinstein, M. A., 1971. The uses of privacy in the good life. *Privacy: Nomos XIII*, 94.

Wenger, E., 1998. *Communities of practice : learning, meaning, and identity.*. Learning in doing: social, cognitive, and computational perspectives, Cambridge : Cambridge University Press, 1998.

Westin, A. F., 1966. Science, privacy, and freedom: Issues and proposals for the 1970’s. part i–the current impact of surveillance on privacy. *Columbia Law Review*, 66 (6), 1003–1050.

Wirtz, B. W. and Birkmeyer, S., 2015. Open government: Origin, development, and conceptual perspectives. *International Journal of Public Administration*, 38 (5), 381 – 396.

Wuyts, K., Scandariato, R. and Joosen, W., 2014. Empirical evaluation of a privacy-focused threat modeling methodology. *The Journal of Systems & Software*, 96, 122 – 138.

Yin, R. K., 2013. *Case study research : design and methods.*. Los Angeles, California : SAGE, 2013.

Yin, R. K., Bateman, P. G. and Moore, G. B., 1985. Case studies and organizational innovation: Strengthening the connection. *Knowledge: Creation, Diffusion, Utilization*, 6 (3), 249–260.

Young, A. and Verhulst, S., 2016. *Open Data Impact when demand and supply meet: key findings of the Open Data Impact Case Studies*. US: NY: The Governance Lab at New York University (the GovLab).

Appendices

Appendix A

FOI Request

Freedom of Information Request submitted to local authorities:

”Dear Sirs,

I would like to request the following information under the Freedom of Information Act:

1. Confirmation that you have a publication scheme under FOI and the web address where this can be found;
2. Contact details (name, job title, email and telephone) of the person(s) responsible for your Freedom of Information process;
3. Confirmation whether you have an open data portal and the web address of where the open source data is published;
4. Contact details of the person(s) (name, job title, email and telephone) who is in charge of your open data portal;
5. Confirmation of whether you submit data to be published on the Apps showcase for OpenDataCommunities;
6. Contact details of the person(s) (name, job title, email and telephone) responsible for submitting/publishing data about on Apps showcase for OpenDataCommunities.

My name is Jane Henriksen-Bulmer and I am a PhD student at Bournemouth University. The subject of my PhD is open data publishing of public body information. The aim is to work in collaboration with public bodies to produce a framework for how public bodies can publish information in open format securely whilst retaining privacy of any personal data.

If you have any queries, please do not hesitate to contact me.

I look forward to hearing from you.

Regards

Jane Henriksen-Bulmer”

A.1 Interview Guide

Local Authority Semi-structured interviews

Rationale for conducting semi-structured interviews

Public bodies are obliged by law to make data available in open format through legislation such as ROPSIR, FOI and INSPIRE. Local Authorities (LA), have been chosen as a good example of a public body to study because of the breadth of data they collect. These datasets will likely include tax collection, housing data, electoral roll, social services and planning applications among others, all datasets that are likely to contain personal or sensitive data and, for that reason, LAs have been chosen as a good example of a public body for these interviews.

The unit of analysis will be the practitioner, in this case the person(s) responsible for deciding which dataset to make available for publishing in open format.

Getting access to and agreement from LA practitioners is likely not going to be easy and therefore, the semi-structured interview technique has been chosen as this method is considered ideal where there is likely to only be one chance to interview a particular participant (Bernard 1988).

The rationale for choosing semi-structured interviews as a method is to develop a comprehensive and sufficient understanding of what policies and processes are in place for open data publishing. The interviews will also try to establish how well developed and integrated these processes are within the local authority (LA) without unduly restricting what information the interviewees (practitioners) may wish to share as part of the interview. Further, the intention is also to understand what data is published, how decisions are made in this regard and the regularity of publication.

Interview Questions

The questions that will be asked are:

Q1 - Team/Department Please describe who is involved in open data publishing with the LA, and how the department/team is set up?

Q2 - Policy What publishing policies and/or guidelines are in place or being developed around Open data with your authority?

Q3 - Process What existing processes are in place for publishing data and are these formal written processes or more loosely defined?

Q4 - Data What information do you currently publish and how are decisions around publication, privacy and the preparation of the data prior to publication made?

Q5 - Problems/obstacles What are the main problems/obstacles that you face in relation to publishing data open source and how often do they occur?

Q6 - Standard Are you aware of the Local Government Classification Scheme (LGCS) and/or the e-GMS standard (e-government Metadata Standard) and, if so, do you currently make use of this resource for publishing information open source?

For validation these questions were tested on two academic staff members with experience in conducting semi-structured interviews. They were also submitted for formal review as part of the University's ethics approval process and approval was granted.

A.2 Participant pre-interview information

The Re-use of Public Sector Information Regulations 2015 require public bodies to publish information open source and where possible the metadata should also be published.

I am interested in the open source publishing process so that I can better understand where and how decisions are made around data, policies and publication. I am particularly interested in how the Re-use of Public Sector Information Regulations 2015 will impact on the current processes.

To give you some idea of the types of things I am interested in finding out about, please find below a list of questions. This is not intended to be an exhaustive list merely to act a guide to the themes/areas that I would like to talk about.

Team/Department I am interested in who is involved in the open data publishing process, and how the department/team is set up.

Policy I would like to discuss what publishing policies and/or guidelines are in place or being developed around open data with your authority.

Process I would like to discuss existing processes for publishing data, whether these are formal written processes or more loosely defined.

Data The regulations state that where possible the raw dataset is to be published. I would like to discuss what information you currently publish and the decisions around publication and the preparation of the data prior to publication.

Problems/obstacles What are the main problems/obstacles that you face in relation to publishing data in open format and how often do they occur?

Standard Are you aware of the Local Government Classification Scheme (LGCS) and, if so, do you currently make use of this resource for publishing data in open format?

Are you aware of the e-GMS standard (e-government Metadata Standard) and, if so, do you currently make use of this resource for publishing data in open format?

A.3 Participant Agreement Form

Full title of the project	
Name, position and contact details of researcher	
Name, position and contact details of supervisor (if the researcher is a student)	

Question	**Please initial or tick here to confirm agreement**
I have read and understood the participant information sheet for the above research project	
I confirm that I have had the opportunity to ask questions	
I understand that my participation is voluntary	
I understand that I am free to withdraw up to the point where the data are processed and become anonymous, so my identity cannot be determined	
During the task or experiment, I am free to withdraw without giving reason and without there being any negative consequences	
Should I not wish to answer any particular question(s), complete a survey or questionnaire I am free to decline	
I understand that all data will be anonymised and I give permission for members of the research team to use my identifiable information for the purposes of this research project	
I agree to the interview being recorded and understand this recording will be used for transcribing purposes and anonymised direct quotes may be used	
I agree to take part in the above research project	
Name of Participant	
Date and Signature	
Name of Researcher	
Date and Signature	

Table A.1: Participant Agreement Form

This form should be signed and dated by all parties after the participant receives a copy of the participant information sheet and any other written information provided to the participants. A copy of the signed and dated participant agreement form should be kept with the project's main documents which must be kept in a secure location.

A.4 Guidance Documents

Axial codes based on the advice or guidance they provide:

1. Data Publishing - HOW - in this category any guidance that relates to how to publish open data. This category consisted of advice around the following sub-categories:

   (a) Data format, style and definitions - suggestions for things to consider;

   (b) Governance - high-level advice on transparency, accountability, quality etc.;

   (c) Portal - advice pertaining to open data portals(s);

   (d) Security - advice about data security in making data available in open format;

2. Data Publishing - WHAT- this category contains any advice about what information or data to publish. This was sub-dividend into:

   (a) Core reference data - i.e. lists and categorisation of what is made available;

   (b) Core subject data - i.e. suggestions for types of documents and data to make available.

3. Data Publishing - WHEN - this category contains any guidance about timings for publication and/or regularity of publication;

4. Data Publishing - WHERE - this category contains any advice about where to publish the open data;

5. Data Publishing - WHO - this category contains any advice about who should be responsible for publication;

6. Definitions - definitions of any open data related terminology;

7. Open Public Data - this consisted of four sub-categories that provide guidance relating to the expected qualities open public data should have/incorporate as follows:

   (a) Publicly available;

   (b) Free and accessible;

   (c) No barriers

   (d) Open Format.

8. Privacy - this category includes advice on how to deal with privacy and privacy risks. This category did not provide much direct guidance, rather it predominately advised that Data Protection regulations should be adhered to.

No.	Guidance Document	Reason for selection for inclusion
1	Local Government Transparency Code 2015 Department for Communities and Local Government (2015)	Provides an overview for local authorities of the types of information they would be expected to publish under the Local Government Transparency Code.
2	The National Information Infrastructure (NII) Implementation Document (Cabinet Office 2015b)	Document setting out the strategy for a national information infrastructure in response to the Shakespeare Review (Shakespeare 2013). This document is high level only.
3	Public Data Principles(Public Sector Transparency Board 2012)	14 public data principles drawn up by the Public Sector Transparency board to outline public data policy on open data for data.gov.uk.
4	Open Government Data Principles (Open Government Working Group 2007)	Original definition of open government data principles.

5	G8 Open Data Charter (G8 2013)	Original Open Data Charter created at the G8 Summit in Loch Erne in June 2013, a collaborative effort between US, UK, France, Canada, EU, Germany, Japan, Italy and Russia.
6	International Open Data Charter (Open Data Charter 2015)	Detailing 6 Open Data principles for how Governments will adopt and integrate these open data principles within their countries.
7	Tim Berner's Lee, five star scheme (Berners-Lee 2006)	This was chosen as a lot of the official guidelines refer to this as a means of assessing how advanced the public body is in open data publication terms.
8	Open Standards Principles (Cabinet Office 2015a)	Part of the Government's Technology Code of Practice which seeks to create agreed standards for a common IT infrastructure, including publishing data in open format.
9	ICO Anonymisation code (ICO 2012)	offers high-level practical guidance on how to anonymise data.
10	APIs: Using and creating Application Programming Interfaces (Gov.UK 2016a)	advices on how to create APIs for open data publishing.
11	Model Publication Scheme (Information Commissioners Office 2015b)	Offers advise and guidance on how to comply with FOI publication scheme, categories of data to include and suggested types of data to publish within each category.
12	Secretary of State's Code of Practice (datasets) on the discharge of public authorities' functions under Part 1 of the FOI (Ministry of Justice 2013)	Guidance for public bodies on how to meet their obligations for making datasets available in open format.
13	Datasets (sections 11, 19 and 45: Freedom of Information Act (Information Commissioners Office 2015a)	Provides an overview of FOI provisions and the publication scheme associated with releasing data within FOI
14	Freedom of Information Act: Principal Local Authorities: Definition document (Information Commissioners Office 2013)	Provides a definition for local authorities of the types of information they would be expected to publish under the FOI publication scheme.
15	Choosing appropriate formats (Information Commissioners Office 2016a)	Offers advise on suitable format of data for open data publication

16	The Guide to the Re-use of Public Sector Information Regulations 2015 (Information Commissioners Office 2016b)	Guidance to public bodies for publishing data in open format under ROPSIR.
17	Template statement on the Re-use of public sector information (The National Archives 2015)	Advice and guidance on what information to make available under ROPSIR

Table A.2: Existing Guidance Documents

Appendix B

CLIFOD Appendix

B.1 Ethics

Before the trial of the CLIFOD questionnaire, the practitioner will be supplied with the following background information:

”This Case Study will be looking at privacy implication of publishing information open source. In particular, I am interested in the privacy decisions made prior to publishing and how these are made.

I propose that the decisions around whether or not publication open source is appropriate be considered using Helen Nissenbaum’s contextual integrity framework.

This framework is framed in terms not only of the data properties within a specific dataset but also requires consideration to be given to the context surrounding the collection, storing and processing of the data and the people involved.

Nissenbaum describes privacy in relation to information as ”a right to appropriate flow of information”. Thus, to apply this to information and navigate the information flows, CI asks practitioners to consider information flow looking at this from the following perspectives:

- Data - i.e. the individual elements that make up the data, this can be described as the rows and columns within a database, each containing pockets of information (attributes);

- Prevailing context - a description of the current context within which the data is collected, stored and processed;

- Actors - These refers to the organisation, department and the people. The people being; the data subjects (i.e. the people who are the subjects of the data itself); and, those who handle the data (the data controller, the data processor and the data receiver);

- Transmission Principles - how the data is conveyed and shared, i.e. the data flow between the actors.

Each of the above elements will be risk assessed and then an evaluation of the surrounding context will be conducted. This will consider:

- Roles - the capacities in which the actors act in a particular context i.e. their job or social role (e.g. may be job role such as Teacher or social role such as student). These roles are then assigned to
  - Activities - how actors interact with others e.g. teaching the child, being taught;
  - Norms - defines the behaviours, what is or is not acceptable; and the duties, privileges, obligations and prerogatives associated with each role, these may implicit or explicit norms so they may be accepted and understood etiquette or prescribed norms.
  - Values - the contextual teleology (the special purpose or use of something), purposes, goals and ends of the particular situation or setting. Further, as part of this risk assessment, any potential mitigation strategies that could enable a potentially failed dataset to be published if implemented will also need to be considered.

From this an informed decision can be made as to whether a particular dataset can be safely published without compromising privacy."

In addition, practitioners will be supplied with and asked to sign a participant agreement form, a copy of which can be found in Section A.3.

B.2 Validation email sent to reviewers

The following email was sent to each of the three expert reviewers for the CLIFOD questions:

"Thank you very much for agreeing to look over my questions for this Case Study.

As explained, what I am planning to do it to try to apply Helen Nissenbaum's Contextual Integrity (CI) framework to a Case Study in collaboration with a Local Authority in the UK. They have agreed to apply the framework to real data in order to decide whether or not to publish data as open data."

The next sentence varied depending on the expert as follows:

Solicitor Before I send this to them and ask them to apply it however, what I would like to do is have some experts evaluate the questions and to that end, I was hoping you would review the questions from a legal perspective...

Practitioner The idea is that they will answer the questions and, depending on the answers, these will be scored on the basis of privacy risks and that will then inform the publishing decision. However, before I do so, it would be very helpful to have some people who understands the 'business' so to speak review the questions.

Academic What I have done so far is to ask a lawyer and a practitioner to review the questions but I could do with a third and I thought someone who could look at them from a methodological perspective would be perfect....rather than send the spread sheet to you (as I have done with the other two) I could send you my methodology which goes into a bit more detail around my thinking for each question that I have come up with and you can then comment on what you think I might have missed, misinterpreted or failed to include.

Whilst I appreciate that you are not familiar with the framework this is not really necessary, I am not looking for you to evaluate the compatibility with that, Rather, what I would like to do is for you, in your role as information officer to evaluate the questions and to that end, I was hoping you would review the questions from an information officer's perspective and provide your comments on:

1. Any areas/questions that you feel have been omitted;
2. Any areas/questions that may need further elaboration or explanation;
3. The order in which the questions are asked, e.g. you may feel that a particular question is better suited to another step or area, or, that some questions are missing from a particular context;
4. Any areas that do not make sense or which you feel are unclear;
5. Any other comments that spring to mind.

The questions have been devised from reading Nissenbaum's book and I have tried to phrase these in more practical terms. In the book Nissenbaum talks about the questions as part of a discussion and explanation but she does not actually, in my opinion, translate this into practical, usable format that a practitioner can follow easily.

With this spreadsheet what I have tried to do is to tease out the main points of the framework and phrase questions around each subject/area so that these can then be applied in practice. The spreadsheet is in 3 parts (worksheets):

- Explanation which is aimed at collecting facts around the dataset to be published;
- Risk Assessment which is aimed at computing risk assessment scores around the facts and evaluate the context of the data collection and the people (the data subjects, data handlers and the data receivers) and how this might affect privacy risks;
- Decision which is basically the decision as to whether or not to publish and the reasons behind the decision reached.

I have included a column in each part for comments, I would be grateful if you could type your comments in there and return the spreadsheet to me with those sections completed.

For information, I have also attached a copy of the Participant Agreement Form that I plan to send the LA which provides an explanation of the framework itself.

If you have any questions or need any more information, please do not hesitate to ask.

Once again, thank you very much for agreeing to evaluate these questions, I really appreciate your input and assistance.

Kind Regards
Jane Henriksen-Bulmer
Bournemouth University”

B.3 Original Questions sent to Reviewers

B.3.1 Key elements

To make it easier for practitioners, these discussions and questions have been translated into CLIFOD, which presents the meta-model in deeper granularity.

In her book, Nissenbaum explains that the framework has three key elements; explanation, evaluation and prescription (p. 190). These elements provided a logical group of overarching categories that would frame the meta-model into understandable, logical progression steps. Thus, the elements have been translated into steps so that each element will, in CLIFOD, be delineated and aid the step-by-step approach logic. The aim of this was to make CLIFOD user friendly and easy to follow.

Explanation (Phase 1 - Explanation)

Nissenbaum explains that the ’explanation’ element refers to the practice or system to be assessed. These should be assessed in view of any ”context-relative informational norms” that may be breached. This should include an assessment of the key ”actors”, i.e. the people that are/could be affected and their ”roles”, as ”data subjects”; ”data senders” or ”data recipients”. It should also consider the ”attributes”, i.e. the information itself (the data) and how this information is transmitted (”transmission principles”) and whether any changes to these elements potentially violate the existing or proposed new information flow.

As this basically involves getting to grips with and understanding the data, the people who work with or are subjects of the data, and how the data is transmitted, this has been translated into Phase 1 and called; ”Explanation” in CLIFOD.

Evaluation (Phase 2 - Risk Assessment)

This part involves assessing the information in view of pertinent ”values, ends and purposes”. Nissenbaum contends that comparing existing flows of information with the proposed new flows, accounting for any breaches (or potential breaches) of values and comparing these to any potential privacy conflict or threat, will afford practitioners the opportunity to identify and mitigate against these. This will in turn enable practitioners to ”establish the significance of each value in light of its contextual ends and purposes” (p 191).

In the meta-model this has been translated into Phase 2; ”Risk Assessment”. The reason behind this decision is that effectively, what the evaluation is trying to achieve is an assessments of the risks associated with any proposed changes or alterations in the data flow.

Prescription (Phase 3 - Decision)

Prescription involves presenting the findings which will guide the practitioner in whether or not a practice or process poses a potential challenge to privacy. This, it is contended, involves making a decision as to the compatibility or non-compatibility of the information for allowing those changes or alterations in the data flow. therefore, this has been translated into Phase 3; ’decision’.

B.3.2 Decision Heuristics

Once the overarching areas had been developed, the next step is to translate the nine step approach and breaking each step down into practical questions that will be asked.

These nine step decision heuristics have been interpreted, and from this, a set of more specific questions created for CLIFOD. The meaning of each of these steps have been outlined below.

B.3.3 Question 1 (Explanation)

”Describe the new practice in terms of information flows” (Nissenbaum 2010)

In order to understand the information flows however, it is necessary to first understand the data. What is the data about and what are the individual parts that make up the datasets. Then, questions can be asked about the existing information flows, i.e. how the data is currently processed. As this relates to finding out about the data itself, this has been placed in the explanation section.

The following questions were devised to cover this area:

- What is the dataset about?
- What attributes are in the data? - please provide a full list

For each attribute (set) within the dataset, please describe:

- Do any of the attributes contain personal information (identifiers)?
- Describe the attribute, exactly what data is collected?
- How many fields/pockets within the dataset contain this type of information?
- Do any of the attributes contain quasi-identifiable data (identifiable through inference data)?
- Describe the attribute, exactly what data is collected?
- How many fields/pockets within the dataset contain this type of information?
- Do any of the attributes contain individual specific attributes (sensitive data)?
- Describe the attribute, exactly what data is collected?
- How many fields/pockets within the dataset contain this type of information?
- Do any of the attributes contain other non-identifiable information (non-sensitive)?
- Describe the attribute, exactly what data is collected?
- How many fields/pockets within the dataset contain this type of information?

B.3.4 Question 2 (Explanation)

”Identify the prevailing context. Establish context at a familiar level of generality (e.g., health care) and identify potential impacts from contexts nested within it, such as teaching hospital.” (Nissenbaum 2010).

This relates to the organisation, what type of organisation and in what context was the data collected. Again, these are facts about the organisational context so this was placed in Phase 1, Explanation.

The following questions were devised to cover this area:

- How is data currently processed?
- Where does data originate (Department)

- Is this data original to this dataset or was data collected from another department/processor?
- If yes, who/which department does the data originate from?

B.3.5 Question 3 (Explanation)

”Identify information subjects, senders, and recipients” (Nissenbaum 2010).

Again, this relates to facts, this time about the people that handle or are the subjects of the data. Therefore, this was placed in Phase 1, Explanation. The questions relating to these were phrased as follows:

Public Body

- What is the public body that owns the information?
- What roles does the public body have? (e.g. library, local authority, University etc.)
- What is the relationship between the public body and the data subject?

Data Subject

- Who/what is the data subject(s)?
- What is the role of the data subject in relation to the data? (e.g. borrower, employee, citizen)
- If data subject is a person what relationship does this person have to the data processor and/or data controller? (none, co-worker, friend, family member, citizen, professional, client, etc.)
- In what context did the data subject divulge the information?
- Was there a legal obligation on the data subject to divulging the information?
- How does data subject interact with data processor/originator? (e.g. friend, co-worker, professionally, citizen, employment) If multiple, please state all that refer

Data Originator

The person who originally collected the information (sender). This would be a data controller but may be a different data controller to the one who is considering whether or not to publish the data, therefore, this has been included with a slightly different name to distinguish from any subsequent data controller(s) who may be making the decisions around the data):

- Who collected the data? (originator) (if multiple, please answer questions for each originator/controller)
- What is the role of the data originator in relation to the data? (librarian, information officer, employee, third-party partner employee etc.)

Data Processor

- Is data processor also data originator?
- What is the role of the data processor in relation to the data? (librarian, information officer, employee, third-party partner employee etc.)
- How may data processors are/have handling/ed the data?
- In what capacity was the data collected by the data processor(s)?
- What is the data processor(s) position in the organisation?
- How does data processor interact with data subject? (e.g. friend, co-worker, professionally, citizen, employment)
- How does data processor interact with data controller? (e.g. friend, co-worker, professionally, citizen, employment)
- How does data processor interact with data recipient? (e.g. friend, co-worker, professionally, citizen, employment)

Data Controller

- Who is the data controller(s)? (if multiple, please answer questions for each controller)
- What is the role of the data controller in relation to the data? (librarian, information officer, employee, third-party partner employee etc.)
- How does data controller interact with data subject(s)? (e.g. friend, co-worker, professionally, citizen, employment)
- How does data controller interact with data processor? (e.g. friend, co-worker, professionally, citizen, employment)

Data Recipient (user)

- Who is the data recipient(s)?
- What is the role of the data recipient(s) in relation to the data? (librarian, information officer, employee, third-party partner employee, unknown etc.)
- What position does the recipient(s) hold?
- How does data recipient interact with data subject(s)? (e.g. friend, co-worker, professionally, citizen, employment)
- How does data recipient(s) interact with data processor(s)? (e.g. friend, co-worker, professionally, citizen, employment)

B.3.6 Question 4 (Explanation)

”Identify transmission principles” (Nissenbaum 2010).

When Nissenbaum discussed how the data is conveyed between the Actors, she refers to this as the ”transmission principles”. Nissenbaum contends that merely considering the data and the actors is not sufficient, consideration also needs to be given to how the data is transmitted and the context in which that transmission takes place. This context needs to include what restrictions may be in place for the data flow and how this is managed and enforced. However, for the purposes of this study, the transmission principle will be open source publishing and therefore, a presumption that no restrictions on the recipients ability to re-use, process and manipulate the data once received will apply. What will be considered will be the format and type of data (dynamic or static) and how often it is proposed this is updated. Again, this relates to factual information about the data and how this is conveyed and therefore, this has been placed in the explanation section.
The following questions will be asked in this category:

- How will the data be shared/published?
- In what format will the data be shared?
- Will data published be static or dynamic?
- If static, how often will it be updated?
- When updated, will the existing data be replaced or will version control or dates be used to denote updates?
- If data published is dynamic will it be real-time?
- Does data contain images or video?

B.3.7 Question 5 (Explanation)

”Locate applicable entrenched informational norms and identify significant points of departure” (Nissenbaum 2010).

This was interpreted as a requirement to consider, not just the social and data etiquette, but also any legal and ethical considerations that need to be taken into account in order to establish the prevailing context. Arguably, this could relate to either explanation or risk assessment as it relates partly to factual information about the data and the purpose of collecting it and partly to the surrounding circumstances which will form part of the risk considerations that will need to be considered. However, a decision was taken that this category is still primarily concerned with gathering facts about the data and thus, this was placed in Phase 1, Explanation. The following questions were devised to capture this category:

- What was the original purpose of the data collection/processing?
- What was the social context of the data collection? e.g. school would be educational context, council tax would be for tax collection etc.
- Was permission/consent sought for processing of the data from data subject(s)?
- Was the data collected with a view to process beyond its original purpose?
- If yes, what was that purpose?
- Was data collected direct from the data subject(s)? (i.e. did they provide the information)
- Was consent sought for original collection/processing purpose?
- If yes, was consent granted for secondary processing purpose?
- If yes, was consent granted for specific secondary processing purpose?
- If yes, what was that purpose?
- Were there any limitations on secondary purpose to which consent was given?
- If yes, what were those limitations?
- Was consent granted for open re-use/sharing/processing?
- Are there any overriding considerations as to why secondary processing should be allowed despite lack of consent/limited consent?
- If yes, what are those considerations?
- Do any of these consideration have legal authority?

B.3.8 Question 6 (Risk Assessment)

”Prima facie assessment: There may be various ways a system or practice defines entrenched norms. One common source is a discrepancy in one or more of the key parameters. Another is that the existing normative structure for the context in question might be ”incomplete” in relation to the activities in question.... A breach of informational norms yields a prima facie judgment that contextual integrity has been violated because presumption favours the entrenched practice” (Nissenbaum 2010).

This was interpreted as an initial evaluation of the facts gathered in Phase 1, in that it asks for thought to be given as to whether those facts in themselves have identified any potential risks or indeed, whether the context surrounding the collection of the data pose a privacy risk.

Therefore, this will be captured through gathering and scoring each sub-category (data; information flows; actors; transmission principles and context) so as to obtain an overview of the dataset and its privacy implications. Further, for each sub-category, risks and mitigation strategies were identified and listed. Finally, practitioners will be given an opportunity to amend the initial score if they decide to apply the identified mitigation and thus, reduce the risk.

B.3.9 Question 7 (Risk Assessment)

”Evaluation I: Consider moral and political factors affected by the practice in question. What might be the harms, the threats to autonomy and freedom? What might be the effects on power structures, implications for justice, fairness, equality, social hierarchy, democracy, and so on? In some instances the results may overwhelmingly favor either accepting or rejecting the system or practice under study; in most of the controversial cases an array of actors emerge requiring further consideration” (Nissenbaum 2010).

This, in CLIFOD, falls into the risk assessment section. It has been placed here as practitioners are asked to start considering the decisions reached in question 6 further. This section refers to the broader context by asking practitioners to now also take into account individual rights, values and norms as well as the surrounding social, political and moral contexts. Thus, here, the following questions will be asked:

Assess disclosure risk

- Disclosure risks identified above?
- Identify any disclosure control processes that may be relevant?

Roles

- Is data subject(s) aware of data being published?
- Has data subject(s) consented to disclosure?
- Has data controller(s) consented to disclosure?

Attributes

- What would be the effect on the attributes published if these are linked to external data, would that pose a new risk?

Values

- Would publication infringe on any political values?
- Would publication impose power imbalance and thus, infringe on any moral values?
- Would publication infringe on any social values?
- Would publication infringe on any moral values?

Norms

- Would publication pose a threat to the autonomy or freedom of the data subject?
- Are there any belief systems that may adversely be affected by publication?
- Would publication result in any form of discrimination?
- Would publication result in informational harm on the data subject?
- Would publication result in breach of confidentiality?
- Would publication result in breach of trust?
- Would publication infringe on any legal compliance?
- Would publication impose any security risks?

Question 8 (Risk Assessment)

”Evaluation II: Ask how the system or practices directly impinge on values, goals, and ends of the context. In addition, consider the meaning or significance of moral and political factors in light of contextual values, ends, purposes, and goals. In other words, what do harms, or threats to autonomy and freedom, or perturbations in power structures and justice mean in relation to this context?” (Nissenbaum 2010).

Here Nissenbaum asks that consideration be given to what potential consequences there might be if any of the Risks, roles or values identified in Question 7 were to occur. This may be either negative or positive. Thus, this forms part of the risk assessment and, as such, has been placed in Phase 2, Risk Assessment. The questions here follow on and elaborate on the context and potential harm identified through the questions asked in the previous sections:

- Is there a reasonable expectation on the part of the data subject and/or data controller of data being kept confidential and not published?
- Are there any privileges or prerogatives that arise from processing or publishing the data from which the public body, the data processor, controller or originator may benefit or be seen to benefit as a result of publication?
- What are the positive values that publication will bring/enhance? These may include commercial gain, improved transparency, meeting legal obligation etc.
- Are there any overriding legal, moral or ethical reasons why publication should be allowed

Question 9 (Decision)

”On the basis of these findings, contextual integrity recommends in favor of or against systems or practices under study. (In rare circumstances, there might be cases that a re-sustained in spite of these findings, accepting resulting threats to the continuing existence of the context itself as a viable social unit” (Nissenbaum 2010).

The final question in Nissenbaum’s nine steps relate to making a decision, thus, this has been placed in Phase 3, Decision.

In CLIFOD, this decision will be to publish or not to publish. Therefore, the questions have been designed to explain the decision taken, what further actions may need to be carried out prior to publication, who is responsible and who will be accountable. Further, if the decision is not to publish, the meta-model asks practitioners to explain the decision so that this may be used to defend or explain the reasoning, in case the information is requested or the decision challenged in future:

- In light of the findings what decision has been reached?

Decision to publish:

- If decision is to publish what work will need to be undertaken prior to publication if any?
- If decision is to publish part of the data rather than the raw data, please specify reason for this decision?
- What is the timescale for publication?
- What is the timescale for publication?
- What processes have/will be put in place for updating of the information?
- Who will be responsible for publication?
- Who will need to be notified of publication?
- Who will be answerable if any questions or challenges arise as a result of publication?

If decision is not to publish:

- What is the reason for non-publication? please provide a full statement including any legal reference where legal constraints form part of the reasoning. This statement should be copiable so that this may be used by data controllers and processors in defence of refusing any request for the release of the data

Appendix C

DPIA Case Study Supporting Documents

C.1 Ethics Approval

Ethics approval for this case study was sought and granted from the University Ethics Committee, a copy of the approval can be found in Figures C.1 to C.4.

C.2 Pilot Study

The work will commence with a pilot study, designed to assess and validate the quality of the questions to be used in phase 1 for the user stories. According to Yin, an appropriate way to validate is through triangulation (Yin 2013). Therefore, for the user stories, the questions will be validated through consultation with 3 independent parties who will be asked to complete the user stories from the perspective of their work.

The intention with collecting these user stories is to get users to formulate in their own words how they use data as part of their daily work. Therefore, to validate the questions and ensure these are appropriate, non-biased and in line with what we are trying to achieve, the questions asked will be validated. The validation will be twofold, first, 3 academic reviewers with experience in conducting case studies will be asked to comment on the quality of the questions and make any suggestions from improvements. Second, three independent reviewers will be asked to take part in a pilot study, providing their user stories.

C.2.1 Pilot User Stories Spread sheet

The academic reviewers will be supplied with a spread sheet that will consist of one worksheet with the first row being a strap-line that reads: Typical Day Story Board. This is then followed by a section for the user to fill in their role or job title.

1. **Tag**: to elicit a tagline for the story being described;
2. **Communication with/data accessed via**: to elicit details of how the data is accessed or, if in person, who the communication is with;
3. **Physical location of user**: to establish where the story takes place;
4. **Data handling method/Device used**: to establish how the data is collected and the type of device used;
5. **Please make an entry for every time data is handled/processed/discussed during a typical day**: for the main narrative of the story, this has intentionally been left vague so that users will use their own style and use words and phrases come naturally to them. Also, this way, the users are more likely to use terminology most commonly used in the environment under study;
6. **How often does this scenario occur?**: This is intended to collect information of the scenario frequency;

C.2.2 Academic feedback

Having reviewed the questions, the feedback from the academics was that, for clarity, it would be useful to add the following questions to the spread sheet:

1. **How long does this scenario take?**: to establish a timeframe;
2. **How often does this scenario occur?**;
3. **Please indicate how demanding you find this task, please explain your reasons**: to understand the users mindset as the story unfolds;
4. **Does this task interfere with any other tasks?**: to elicit level of disruption (actual or perceived);
5. **Comment**: for the pilot study element only, to allow pilot users to add comments about the experience and/or quality of the questions.

Further, the academics also suggested it may be useful to provide some examples for the users to illustrate what is meant by each question asked and therefore, two general examples were devised around general office work and added to the pilot user spread sheet (see Appendix C, Section C.3).

C.2.3 Pilot User Stories

For the pilot study, three independent participants were recruited. To ensure these participants did not have any links to the Charity or know any of the Charities employees or volunteers, two of the participants were recruited from outside the local area, the third confirmed that they had no links with any charities in the local area that deal with vulnerable people or their staff.

In line with the case study aims, each of the pilot participants will be provided with a participant information sheet (see Appendix C, Section C.4) and asked to complete a participant agreement form (see Appendix A, Section A.3). Upon agreeing to take part in the pilot study, the pilot participants were then asked complete a user story spread sheet, describing their duties or tasks as part of a typical day.

C.2.4 Pilot User Stories - results

The pilot study participants all completed the questionnaires with no problems. None of them raised any issues over the questions or the format of the questions.

Interestingly, while two participants provided details of a few different scenarios of when they collect or handle data as part of their day, the third participant chose to use the spread sheet itself as part of the story. What this participant chose to do was to recount their week and the tasks they handle each day/part day. Then, as part of the story they described what data they handle at each stage, telling the story of their week, detailing what happens to the data they handle as the story is recounted over the week. This would indicate that the aim of having participants use their own words and style was effective.

C.3 Pilot User Studies - examples

These questions were added to the pilot user story spreadsheet.

A further suggestion was made to add an example or two to provide users with some idea of what is expected. This might help elicit more detailed responses.

To this end, the following examples were added to the spread sheet (first two rows completed):

Example 1

1. **Tag**: am - meeting with client;
    2. **Communication with/data accessed via**: Client;
    3. **Physical location of user**: Office;
    4. **Data handling method/Device used**: Manual, paper record and email (computer);

5. **Please make an entry for every time data is handled/processed/discussed during a typical day**: Met with AB, this was our first meeting. We went through the list of orders and agreed which ones needed invoicing. I then processed those records and created an invoice for each client. This involved inputting the client details onto the system together with the list of items ordered and the prices. The system calculates the VAT and totals automatically but I always check to make sure the prices are correct as sometimes, the prices come up wrong. If that happens, I have to get in touch with the order team so they can update the prices before I can create the invoice;
6. **How long does this scenario take?**: The meeting lasted about 2 hours, I then spent another 4 hours creating the invoices;
7. **How often does this scenario occur?**: we invoice once a week
8. **Please indicate how demanding you find this task, please explain your reasons**: This is quite a time consuming task but as long as all the prices are correct, it is not difficult. I do need to be vigilant with the checking though as we don't want to send an incorrect invoice out;
9. **Does this task interfere with any other tasks?**: When I am in a meeting with a client no-one usually disturbs us. The phone is put on silent and there is no computer in the meeting room. However, as soon as the meeting finishes and I get to the admin part, there tends to be constant interruptions.

Example 2

1. **Tag**: - checked emails
2. **Communication with/data accessed via**: Internet;
3. **Physical location of user**: Home;
4. **Data handling method/Device used**: Home computer;
5. **Please make an entry for every time data is handled/processed/discussed during a typical day**: Logged into emails, there were two emails that I needed to respond to. One concerned the system upgrade planned for tomorrow morning, I needed to respond to make sure this does not happen until I have run the backup;
6. **How long does this scenario take?**: About 20 minutes;
7. **How often does this scenario occur?**: Every day
8. **Please indicate how demanding you find this task, please explain your reasons**: The task was not difficult, more disappointing because the IT team could have informed me sooner so I could have done the backup before I left today. Now I have to go in early tomorrow;

9. **Does this task interfere with any other tasks?**: Not really, I do not have to check emails from home but chose to so I don't arrive in the office the next morning to too many surprises.

C.4 Participant Information Sheet - Pilot Study

"First of all, thank you very much for agreeing to take part in this pilot to trial our user story questionnaire, we really value your input, time and comments.

BU has a working relationship with a charity that provides information, advice and support for those vulnerable people. Like many organisations, they struggle to understand how best to implement GDPR, and finding sufficient time and resources to demonstrate compliance with the regulation by May 2018.

We will be collaborating with this charity first of all, to gather information about what data they work with and how this happens by asking them to complete a user story questionnaire. The user stories we want to capture are those that involve any form of data capturing, manipulation, processing or sharing .

The idea is that we will ask each member of staff/volunteer etc. within the organisation to complete their user story for us for any task that user carries out as part of their daily work, whether that is at work, in the field or at home that involves handing information and/or data.

We will then use this information to help us better understand how data is collected, handled and shared both within the organisation, externally and with any third party stakeholders.

However, we are conscious that we want to collect quality information from the participants without being too prescriptive in what we collect, as we want them to use their own words in describing and explaining what they do as part of their job.

Therefore, we have asked you to help us by trialling a pilot version of our questionnaire and then providing some feedback for us.

To explain, we have provided a couple of examples of instances when an invoice clerk might handle data.

What we would like you to do is to complete the attached user story questionnaire for your role or job by describing the times during your typical working day that you handle data for us. Please complete a row for each occurrence.

Then, we would also like you to comment on the questions that we ask and whether these make sense to you. If you feel there is something that needs explaining further or requires re-phrasing to make sense. A comment section has been provided at the end of the questionnaire for you to write your observations and feedback.

Once again, thank you for agreeing to participate, we really appreciate your help."

C.5 Participant Information Sheet - Charity user stories

The Participant information sheet for staff and volunteers, states:
”The Charity”, like many organisations, struggle to understand how best to implement GDPR, and finding sufficient time and resources to demonstrate compliance with the regulation by May 2018.

To help with this, BU are working with ”the Charity” to assist with GDPR implementation and, as part of this, Jane Henriksen Bulmer, a BU Research Assistant, is conducting a case study that will help ”the Charity” implement GDPR and demonstrate compliance.

For this study, first of all, we need to gather information about what data the charity works with and how this happens. To this end, we would like to start by gaining an understanding of how you, as a staff member, use information as part of your daily work by completing a user story questionnaire (attached).

The user stories we want to capture are those that involve you handling any form of data or information. That may be as part of a conversation, in writing or capturing, manipulation, processing or sharing information electronically.

The idea is that we will ask each member of staff/volunteer etc. within the organisation to complete their user story for us for any task that the user carries out as part of their daily work, whether that is at work, in the field or at home that involves handing information and/or data.

We will then use this information to help us better understand how data is collected, handled and shared both within the organisation, externally and with any third party stakeholders.

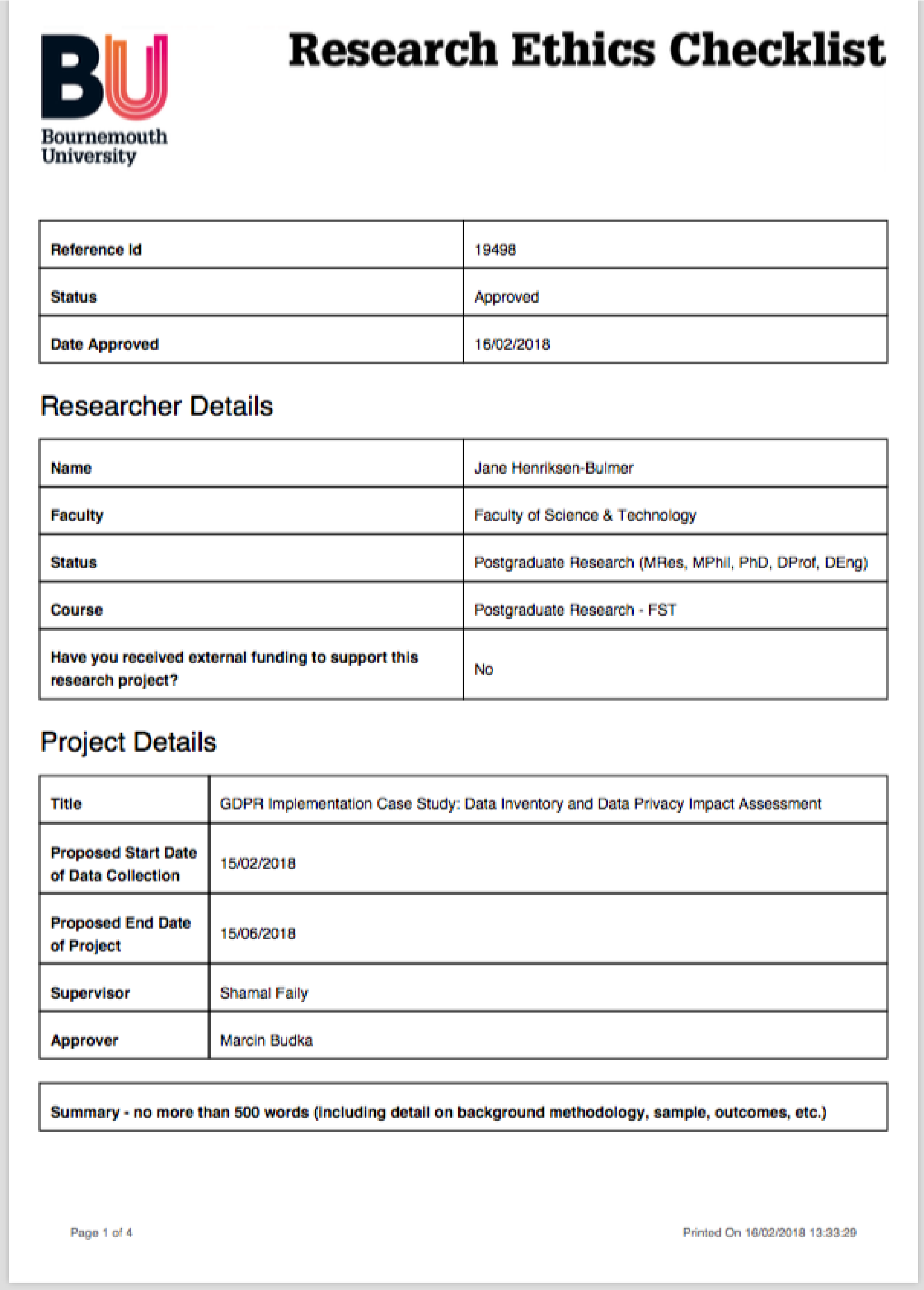
To explain, we have provided a couple of examples of instances when a fictitious case worker might handle data. We appreciate that there may be many more instances in a typical working day but these are only intended as examples.

What we would like you to do is to complete the attached user story questionnaire for your role or job by describing the times during your typical working day that you handle data for us. We would like you to tell the story of what happens/what you do with the form/data as part of the scenario. Please complete a row for each occurrence or scenario.

If you have any questions or queries, please feel free to either speak with your supervisor or contact Jane Henriksen-Bulmer direct via email on jhenriksenbulmer@bournemouth.ac.uk

For research and ethical purposes, we also need for you to complete a participant sheet which explains how the data we collect will be used as part of our research. This form is at the end of this document. Please complete and sign this form and return it to us together with the completed user story questionnaire.

Finally, we would like to take this opportunity to thank you very much for assisting us with this work, we really value your input, time and comments.”



Bournemouth University

Research Ethics Checklist

Reference Id	19498
Status	Approved
Date Approved	16/02/2018

Researcher Details

Name	Jane Henriksen-Bulmer
Faculty	Faculty of Science & Technology
Status	Postgraduate Research (MRes, MPhil, PhD, DProf, DEng)
Course	Postgraduate Research - FST
Have you received external funding to support this research project?	No

Project Details

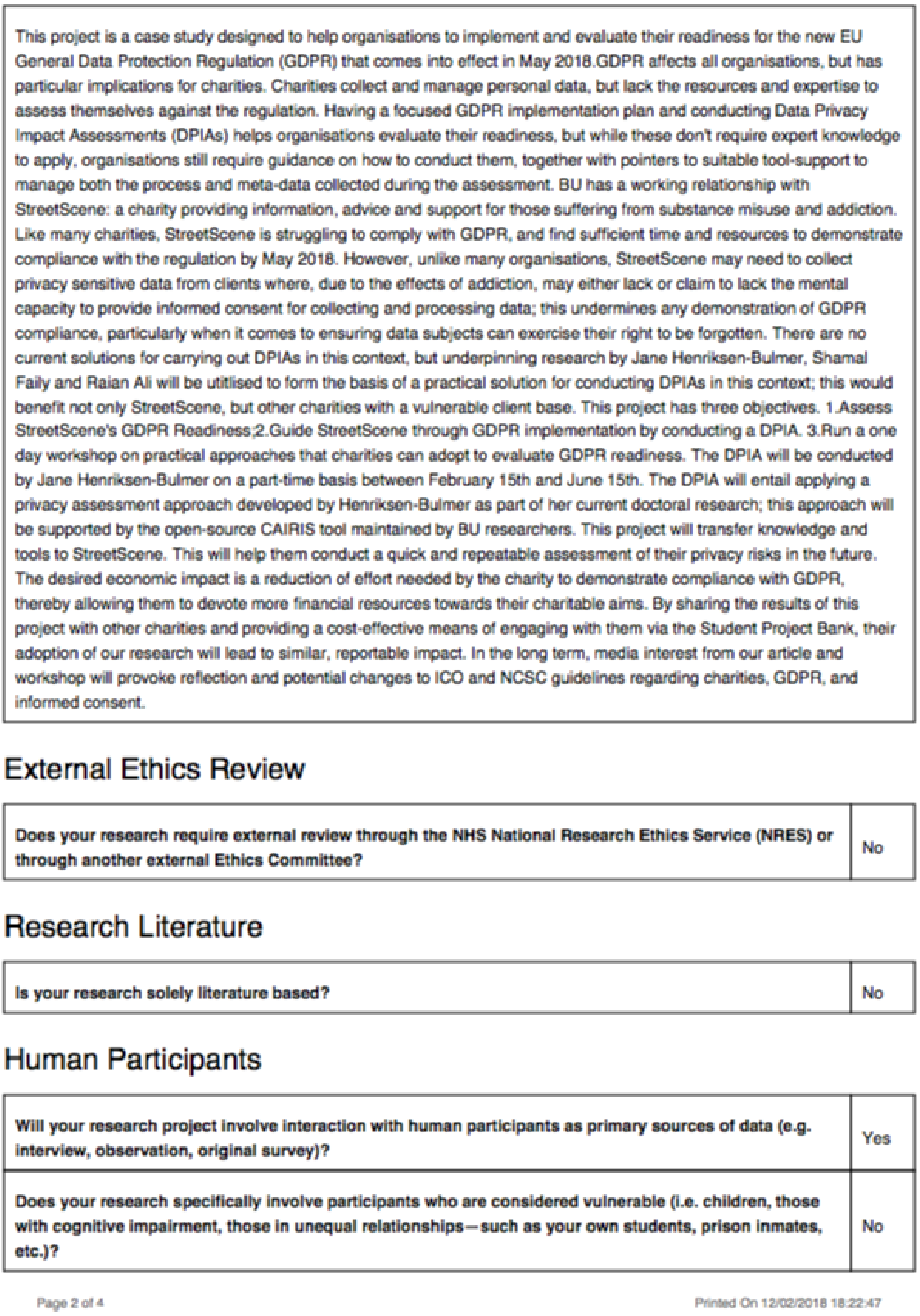
Title	GDPR Implementation Case Study: Data Inventory and Data Privacy Impact Assessment
Proposed Start Date of Data Collection	15/02/2018
Proposed End Date of Project	15/06/2018
Supervisor	Shamal Faily
Approver	Marcin Budka

Summary - no more than 500 words (including detail on background methodology, sample, outcomes, etc.)

Page 1 of 4

Printed On 16/02/2018 13:33:29

Figure C.1: Ethics Approval Form - Page 1



This project is a case study designed to help organisations to implement and evaluate their readiness for the new EU General Data Protection Regulation (GDPR) that comes into effect in May 2018.GDPR affects all organisations, but has particular implications for charities. Charities collect and manage personal data, but lack the resources and expertise to assess themselves against the regulation. Having a focused GDPR implementation plan and conducting Data Privacy Impact Assessments (DPIAs) helps organisations evaluate their readiness, but while these don't require expert knowledge to apply, organisations still require guidance on how to conduct them, together with pointers to suitable tool-support to manage both the process and meta-data collected during the assessment. BU has a working relationship with StreetScene: a charity providing information, advice and support for those suffering from substance misuse and addiction. Like many charities, StreetScene is struggling to comply with GDPR, and find sufficient time and resources to demonstrate compliance with the regulation by May 2018. However, unlike many organisations, StreetScene may need to collect privacy sensitive data from clients where, due to the effects of addiction, may either lack or claim to lack the mental capacity to provide informed consent for collecting and processing data; this undermines any demonstration of GDPR compliance, particularly when it comes to ensuring data subjects can exercise their right to be forgotten. There are no current solutions for carrying out DPIAs in this context, but underpinning research by Jane Henriksen-Bulmer, Shamal Faily and Raian Ali will be utitlised to form the basis of a practical solution for conducting DPIAs in this context; this would benefit not only StreetScene, but other charities with a vulnerable client base. This project has three objectives. 1.Assess StreetScene's GDPR Readiness;2.Guide StreetScene through GDPR implementation by conducting a DPIA. 3.Run a one day workshop on practical approaches that charities can adopt to evaluate GDPR readiness. The DPIA will be conducted by Jane Henriksen-Bulmer on a part-time basis between February 15th and June 15th. The DPIA will entail applying a privacy assessment approach developed by Henriksen-Bulmer as part of her current doctoral research; this approach will be supported by the open-source CAIRIS tool maintained by BU researchers. This project will transfer knowledge and tools to StreetScene. This will help them conduct a quick and repeatable assessment of their privacy risks in the future. The desired economic impact is a reduction of effort needed by the charity to demonstrate compliance with GDPR, thereby allowing them to devote more financial resources towards their charitable aims. By sharing the results of this project with other charities and providing a cost-effective means of engaging with them via the Student Project Bank, their adoption of our research will lead to similar, reportable impact. In the long term, media interest from our article and workshop will provoke reflection and potential changes to ICO and NCSC guidelines regarding charities, GDPR, and informed consent.

External Ethics Review

Does your research require external review through the NHS National Research Ethics Service (NRES) or through another external Ethics Committee?	No

Research Literature

Is your research solely literature based?	No

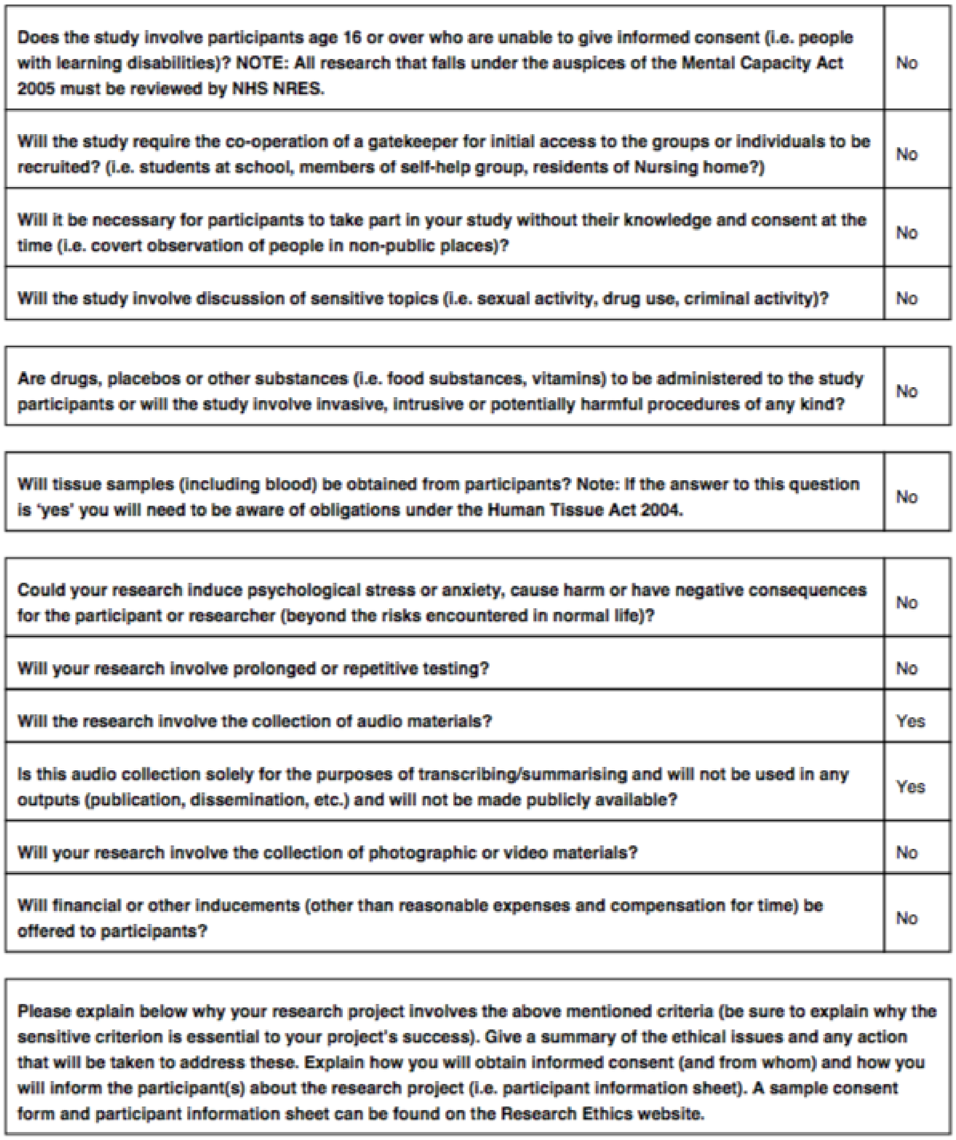
Human Participants

Will your research project involve interaction with human participants as primary sources of data (e.g. interview, observation, original survey)?	Yes
Does your research specifically involve participants who are considered vulnerable (i.e. children, those with cognitive impairment, those in unequal relationships—such as your own students, prison inmates, etc.)?	No

Page 2 of 4

Printed On 12/02/2018 18:22:47

Figure C.2: Ethics Approval Form - Page 2



Does the study involve participants age 16 or over who are unable to give informed consent (i.e. people with learning disabilities)? NOTE: All research that falls under the auspices of the Mental Capacity Act 2005 must be reviewed by NHS NRES.	No
Will the study require the co-operation of a gatekeeper for initial access to the groups or individuals to be recruited? (i.e. students at school, members of self-help group, residents of Nursing home?)	No
Will it be necessary for participants to take part in your study without their knowledge and consent at the time (i.e. covert observation of people in non-public places)?	No
Will the study involve discussion of sensitive topics (i.e. sexual activity, drug use, criminal activity)?	No

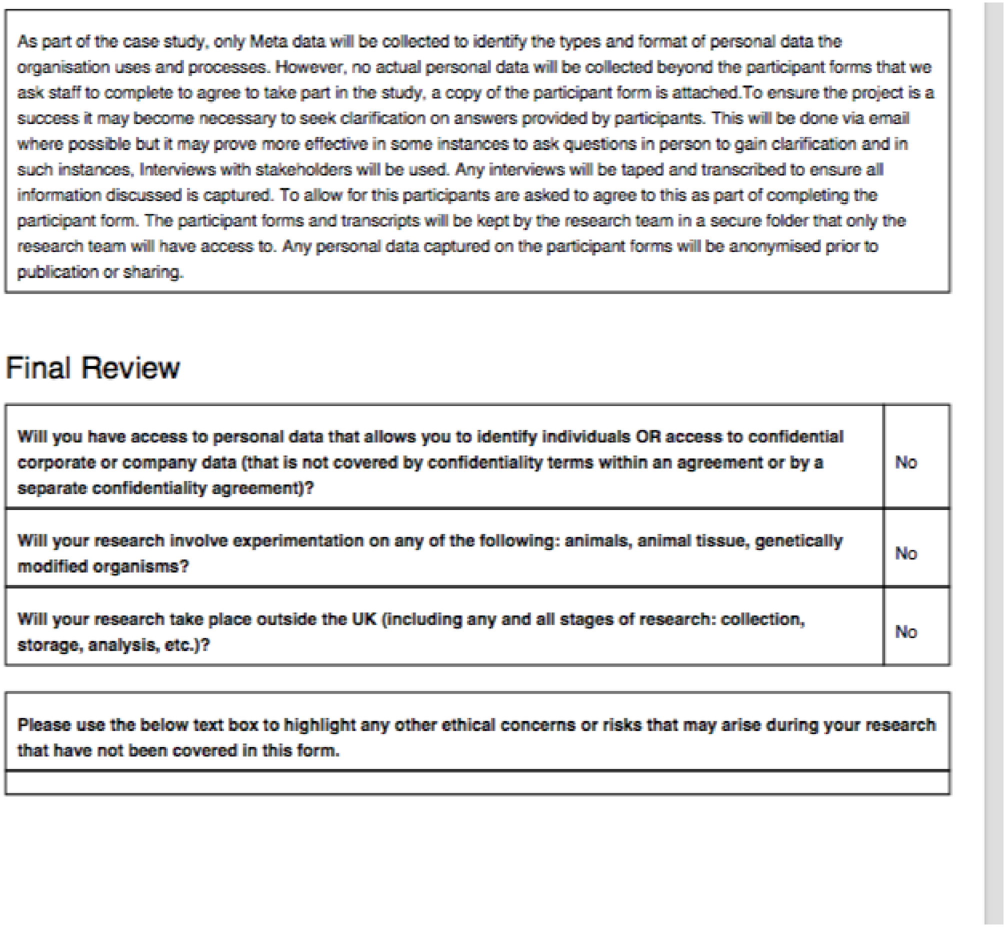
Are drugs, placebos or other substances (i.e. food substances, vitamins) to be administered to the study participants or will the study involve invasive, intrusive or potentially harmful procedures of any kind?	No

Will tissue samples (including blood) be obtained from participants? Note: If the answer to this question is 'yes' you will need to be aware of obligations under the Human Tissue Act 2004.	No

Could your research induce psychological stress or anxiety, cause harm or have negative consequences for the participant or researcher (beyond the risks encountered in normal life)?	No
Will your research involve prolonged or repetitive testing?	No
Will the research involve the collection of audio materials?	Yes
Is this audio collection solely for the purposes of transcribing/summarising and will not be used in any outputs (publication, dissemination, etc.) and will not be made publicly available?	Yes
Will your research involve the collection of photographic or video materials?	No
Will financial or other inducements (other than reasonable expenses and compensation for time) be offered to participants?	No

Please explain below why your research project involves the above mentioned criteria (be sure to explain why the sensitive criterion is essential to your project's success). Give a summary of the ethical issues and any action that will be taken to address these. Explain how you will obtain informed consent (and from whom) and how you will inform the participant(s) about the research project (i.e. participant information sheet). A sample consent form and participant information sheet can be found on the Research Ethics website.

Figure C.3: Ethics Approval Form - Page 3



As part of the case study, only Meta data will be collected to identify the types and format of personal data the organisation uses and processes. However, no actual personal data will be collected beyond the participant forms that we ask staff to complete to agree to take part in the study, a copy of the participant form is attached.To ensure the project is a success it may become necessary to seek clarification on answers provided by participants. This will be done via email where possible but it may prove more effective in some instances to ask questions in person to gain clarification and in such instances, Interviews with stakeholders will be used. Any interviews will be taped and transcribed to ensure all information discussed is captured. To allow for this participants are asked to agree to this as part of completing the participant form. The participant forms and transcripts will be kept by the research team in a secure folder that only the research team will have access to. Any personal data captured on the participant forms will be anonymised prior to publication or sharing.

Final Review

Will you have access to personal data that allows you to identify individuals OR access to confidential corporate or company data (that is not covered by confidentiality terms within an agreement or by a separate confidentiality agreement)?	No
Will your research involve experimentation on any of the following: animals, animal tissue, genetically modified organisms?	No
Will your research take place outside the UK (including any and all stages of research: collection, storage, analysis, etc.)?	No

Please use the below text box to highlight any other ethical concerns or risks that may arise during your research that have not been covered in this form.

Figure C.4: Ethics Approval Form - Page 4

Appendix D

DPIA Case Study - Supporting documents

D.1 From GDPR & ICO to the DPIA Data Wheel

This work started by looking at what the DPIA is actually trying to achieve, a privacy assessment of data as processing activities during its lifecycle with the organisation, and how this could be adapted to fall into the CI phases devised in previous work *explanation, risk assessment and decision*. Research shows that modelling is an effective tool for communication of a cause of action, a programme or a process. For example, visualisation has been shown to help explain ”how to get there” (Millar et al. 2001); and enhance understanding and communication, e.g. of risks (Lipkus and Hollands 1999). Therefore, a logic model was created to try to fit the process into some form of acronym or model that would provide a visual representation that practitioners could use for a quick overview.

The analysis began by taking the 9 steps for conducting a DPIA outlined by the ICO (ICO 2018a) and mapping these to how they relate to the 6 GDPR principles (see Chapter 7, Section 2.4.4). This was done through a series of logic models, starting with in Figure D.1, with the first column depicting ICO principles and the corresponding GDPR principle showing in the second column.

From this, the analysis started to look at how each of these stages can be described in questions that cover both CI and the GDPR DPIA. To this end, a process of working the concepts began. This is depicted through a series of logic models that map these concepts and how they releat to the three phases of contextual integrity (CI) as used in the meta-model and CLIFOD: *Explanation, Risk Assessment and Decision* (see Chapters 5 and 6). This process is depicted in Figures D.2, D.3 and D.4.

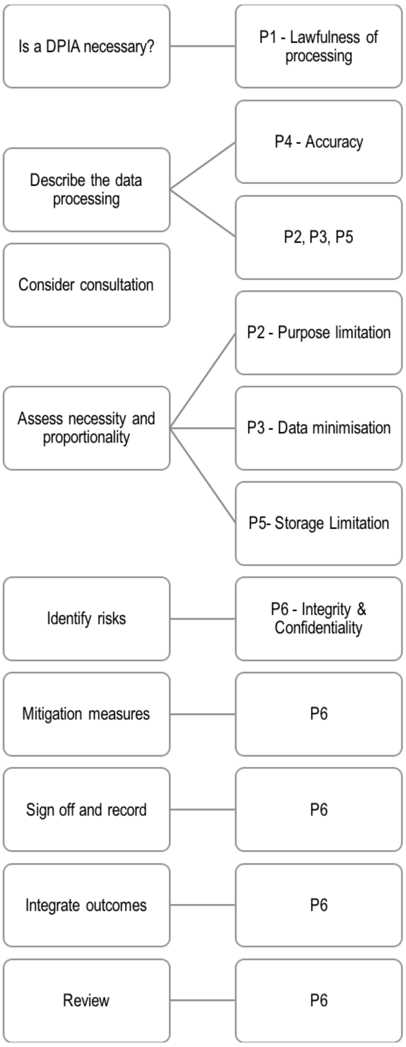


Figure D.1: Aligning GDPR to ICO DPIA Guidance

D.2 Reviewed Privacy Policy Headings

D.2.1 Data Subject Privacy Policy

Introduction providing an overview stating the Charity's commitment to protecting personal data and the rights of individuals together with a brief outline of the rest of the privacy policy;

Personal data that we collect imparting details about the what personal data the Charity collects;

How we collect your personal data detailing how the data is collected;

How we use your personal data i.e. details about how the data is used (processed);

Who we share your personal data with details about who the data may be shared with;

How we safeguard your personal data i.e. how the data is safeguarded and protected;

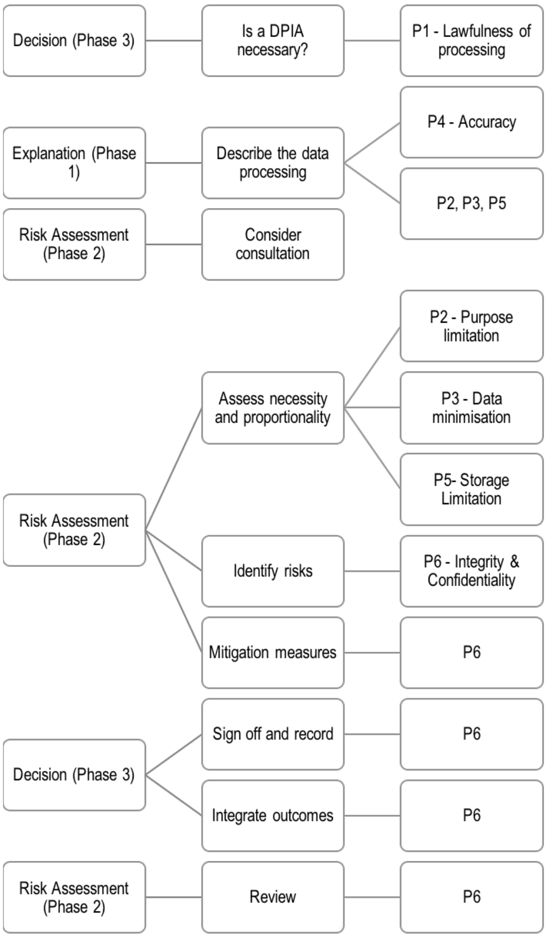


Figure D.2: Aligning GDPR & ICO DPIA Guidance to Contextual Integrity

How long we keep your personal data i.e. details pertaining to how long the data is kept and how it is disposed of at the end of its lifecycle;

Your Rights relating to your personal data providing detailed information about what rights the individual's (the data subjects) in respect of their data (Right of: access, rectification, erasure, to restrict processing, notification, data portability and objection) and how they may invoke these rights;

Examples of the types of Personal Data we collect i.e. a list of examples of the types of data collected.

D.2.2 Staff Privacy Policy

The privacy policy for staff was also updated to ensure staff are fully aware of their obligations under GDPR. The new document devised covers both procedure and obligations under GDPR and includes the following sections:

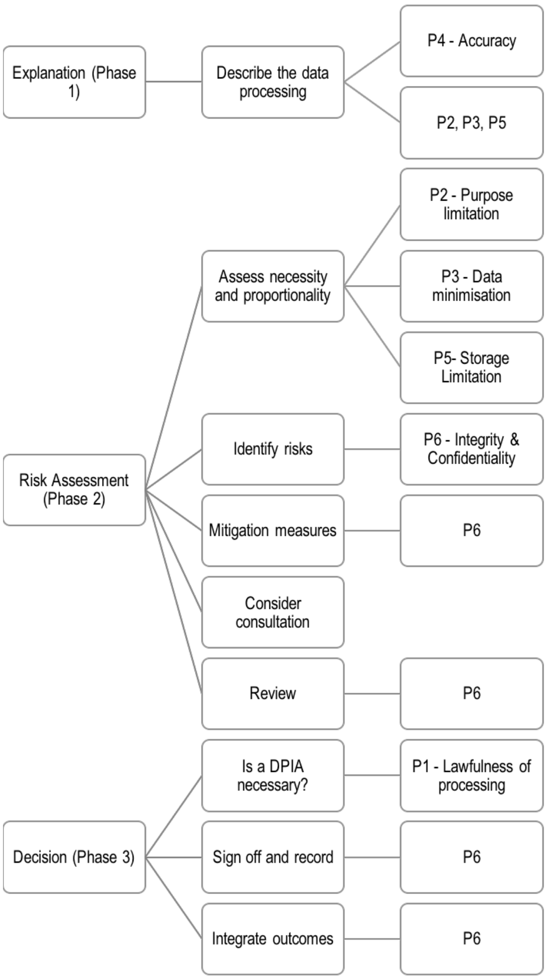


Figure D.3: Aligning GDPR & ICO DPIA Guidance to Contextual Integrity

Policy Statement explaining that all data must remain confidential and, as such, needs to be treated with care and in accordance with the law;

Summary of Principles a brief outline overview of the GDPR principles;

Requirements of the Regulation informing staff they must adhere to GDPR when handling personal data;

Responsibilities of senior staff outlining expectations of management staff in overseeing compliance with GDPR and that procedures laid down within the policy adhered to;

Responsibilities of all staff setting a requirement for all staff to be aware of and comply with GDPR principles and guidelines;

Data Security information staff of their responsibilities when handling personal data to keep such data safe and secure;

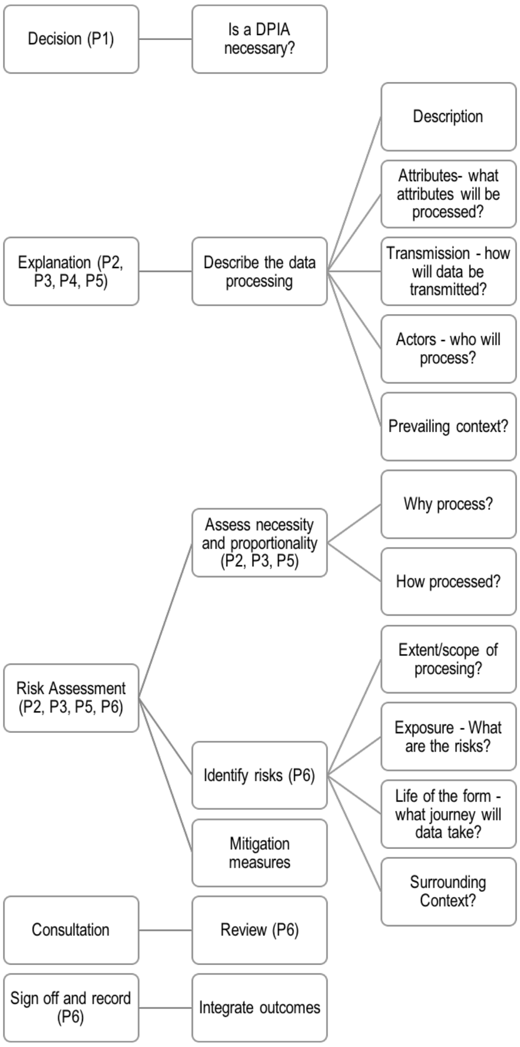

Decision (P1)
Is a DPIA necessary?
Description
Attributes- what attributes will be processed?
Explanation (P2, P3, P4, P5)
Describe the data processing
Transmission - how will data be transmitted?
Actors - who will process?
Prevailing context?
Why process?
Assess necessity and proportionality (P2, P3, P5)
How processed?
Extent/scope of procesing?
Risk Assessment (P2, P3, P5, P6)
Exposure - What are the risks?
Identify risks (P6)
Life of the form - what journey will data take?
Mitigation measures
Surrounding Context?
Consultation
Review (P6)
Sign off and record (P6)
Integrate outcomes

Figure D.4: Aligning GDPR & ICO DPIA Guidance to Contextual Integrity

Data Handling Procedures for :

i Report Writing and transferring of sensitive information: placing an obligation on staff to pseudonymise data where possible and to encrypt and/or password protect data prior to transmitting/sharing;

ii USB Keys;

iii Home computers/mobile phones/Tablets;

iv Work computers;

v Client Diaries/Assignments;

vi Paper with client's name;

vii Emails regarding a client;

viii What to do if you become or are aware of a Data Breach;

ix Subject Access Requests;

x Data Protection Officer.

D.3 Client Records

D.3.1 Client Assessment

This consists of:

1. Front cover (index): mainly used for admin purposes, this captures whether bed was offered, the date of entering treatment, funding arrangements etc.;
2. Personal details: where the clients personal details are captured;
3. Consent: capturing granular consent from the client about what personal details may or may not be shared with whom and under what circumstances;
4. Professionals: this captures details of the various professional bodies and individuals who are or have been involved in providing treatment and advice to the client prior to entering the Charity for treatment;
5. Health assessment: this is an overview of the clients current and past health history;
6. Psychiatric history: this captures details about the client's psychiatric history and any previous treatment undertaken;
7. Brief using history: a recount of the client's history of using substances and/or alcohol;
8. Legal/Forensic: collects details of offences and any criminal convictions
9. Social Assessment: designed to capture the social context, details of family, any dependents and social history of the client;
10. Finance: details of client's financial situation;
11. Risk Assessment: an assessment of risks to the client while in residence for treatment;
12. Agreement: client's agreement that details provided within the assessment are correct and consent to enter treatment.

D.3.2 Care Plan

This consists of:

1. Incident Log: a log of any incidents that have occurred while client in active treatment (e.g. contracts or breach of house rules & expectations);
2. Risk Assessment (e.g. risk of relapse, self-injury / suicide / depression etc.)
3. Treatment Goals
4. Treatment Goals - Workshops
5. Agreed Interventions
6. Social Goals
7. Personal Goals
8. Cultural
9. Relapse Prevention
10. Harm Minimisation
11. Diet
12. Exercise
13. Phases
14. Education
15. Restrictions
16. Finances
17. Housing Support
18. Housing Support

IN ADDITION, THE CARE PLAN ALSO CONTAINS DETAILS/ASSESSMENTS ON:

- Alcohol detox
- Multi-drug & alcohol
- Parenting skills
- Chemical history
- Medication: profile, administration, discharge, error, record of variable doses, transfer

- Drug test results
- Risk assessment
- Treatment plan
- Treatment progress

TYPES OF CARE PLAN:

- First stage
- Continuing
- Follow-on

D.3.3 CLIENT FILE FORMAT/LAYOUT

Documents filed in Client file:
PRE-SECTION:

- Front page contact sheet
- Checklist 12/24 weeks
- Telephone contacts
- Drug testing sheets
- Supervision notes
- Therapeutic agreements

SECTION ONE:

1. Assessment
2. Client details sheet
3. Expectation of the community
4. Control of substances hazardous to health (C.O.S.H.H) form
5. Total abstinence agreement
6. Client consent form
7. Smoking policy
8. Injury disclaimer

9. Complaints procedure
10. Valuables form
11. Service user guide

SECTION TWO:

- Treatment outcomes profile (TOPS) forms
- Any correspondence in and out
- Financial contract

SECTION THREE:

- 3 Questionnaires:
    - i What I want from treatment
    - ii Relationship assessment
    - iii Dependency questionnaire
- Social History

SECTION FOUR:

- Check-ins
- Progress reports/ Weekly goals
- Care plan/ risk assessment

SECTION FIVE:

- Mini group
- Progress reviews (goals group)
- Weekly plans
- Any other work completed by resident

D.4 Life of the Form analysis

Based on the analysis of the life of the form responses and the staff user stories, the most commonly used forms have been listed in the sections below. For ease, these have been broken down into data relating to clients and data relating to staff and data handling processes.

D.4.1 CLIENT DATA

Starting with the largest section of template forms, those relating to clients, the main forms and data handling processes carried out by staff as part of their daily work in relation to client data were:

Client Assessment (aka Assessment) this is a 5-section paper-based form that clients and staff complete when the client first enters for treatment (see Appendix C, Section D.3 for full list of template forms included in the client assessment);

- The client assessment is completed as part of the initial meeting with the client to assess their suitability for treatment. If accepted and the client wishes to enter into treatment at one of the Charity's houses, the assessment forms part of the contract;
- The client assessment is a living document throughout the client's stay at the Charity which means they are regularly removed from filing to be updated;
- Some data from the assessment is transferred onto the client register (electronic register) of current clients residing at one of the houses;
- Any assessments that have been worked with and/or updated that day, are placed back in the filing cabinet within the office and locked away at the end of the day;
- The assessment will travel with the client if they move to a new house or when they leave treatment if they request their file;
- Parts of the assessment may travel independently or the information on parts of the form may be used for other forms, e.g.
  - i Internal Departments: Details of professional support workers who work with the client or medication records may be transcribed onto other forms and shared with internal stakeholders, such as medical information being transferred onto forms in the project office forms;
  - ii External Stakeholders: Details from the client assessment may be transcribed onto other forms and shared with external stakeholders, such as one of the client's professional support workers, the referrer or funder or medical data being shared with the client's doctor. For this type of sharing, the Charity does not rely on contract as the lawful basis for processing, granular consent is also sought from the client for sharing with external stakeholders;
  - iii Invoicing: Personal details and finance details shared with finance department (internal) so they can raise invoices to charge for treatment. It is not clear how much personal data about the client will appear on invoices to funder/referrer.

- The final journey the assessment will take is to travel into storage once the client's treatment is complete and be stored for 7 years. The admin department oversee and manage the storing of records and ensures that all records are securely shredded once the seven years have passed.

Client Care Plan (aka care plan or treatment plan) :

- There are a number of different care plans for different stages of treatment. These consist of a record of agreed interventions, any incidents while the client is in treatment and a series of goals (aims and/or needs) that the client seeks to achieve from entering treatment. These include, personal, social, cultural, religious, treatment and relapse prevention goals as well as an ongoing risk assessment of client's state of mind and any potential obstacles to progress (see Appendix C, Section D.3 for full list of template forms included in the care plan);
- The care plan(s) are paper based;
- The care plan(s) are large living documents used throughout the client's treatment and as part of any out-patient ongoing treatment after the client has completed in-house treatment;
- The care plan is regularly updated with details of the client's condition and treatment;
- Part of the information contained within the care plan may be shared with third parties as part of treatment e.g. the care manager or doctor.

Medication Records (a.k.a. MAR sheets) MAR records of client's medication administration while in treatment.

- MAR sheets for each client are kept with the medicines in a locked cabinet. These are updated each day with details of what medication the client has been issued with etc. and returned to the locked cabinet;
- Once complete, MAR sheets are filed and stored with the client assessment file;
- Part of the information from the MAR sheet may be used to re-order prescription medication from the Client's doctor. This order may be emailed or faxed to the doctor for approval and/or the chemist to place the order.

Client Register/Register of Clients register of clients used for internal admin purposes only.

- The client register contains personal details about all current clients, their next of kin and professional support workers;

- These details are kept in an electronic spreadsheet, accessible by all staff but not shared outside the organisation;
- The spread sheet may be shared with internal stakeholders via email;
- The spread sheet is password protected. It is not clear how regular the password is updated or how updated passwords are communicated to stakeholders.

Reports there are a number of different reports produced, some for internal purposes, others are external:

- Internal: the following reports are internal reports, although statistics based on the information may be produced from these which may be shared. Where that is the case the data is pseudonymised or anonymised and only statistics are published:
  i Bods in beds i.e. how many beds are occupied within the organisation at any one time; and
  ii Evaluation reports based on client feedback. Unfortunately, it was not clear from the data whether or not these feedback forms are kept, where or for what time period the data is kept;
- External: a number of reports are produced and to external bodies for statistical purposes e.g. data is shared with the National Drug Treatment Monitoring System (NDTMS) and other safeguarding and care standards, such as the Quality Assessments Framework (QAF) are also completed and shared. However, some of these still contain some personal information, e.g. some of the NDTMS reports that are sent to NDEC contain client names, location and dates of birth.

Office notice board in some houses a notice board is on display in the office with details of client appointments and commitments for the week. Although the office is locked, anyone who enters (clients or guests) will be able to see the notice board.

D.4.2 STAFF DATA

In relation to staff data, the following forms and/or processes were mentioned/discussed as part of the staff stories and life of the form analysis:

Staff Diary there is a central diary electronically, accessible by all staff. This contains details of appointments etc. that clients must keep. There is also, in some instances a manual central diary, kept in the office;

Staff Rota this contains staff first names and when they are due to work. The Rota is created and stored electronically. The Rota is updated by managers each day and a copy is emailed to:

- all staff and a hard copy is printed and posted on the wall in the office and shredded once time period covered has passed;
- HR to ensure all shifts are covered;
- finance to check and ensure staff work are not asked to/do not exceed the working time directive on working hours. It is not clear whether the Rota that is emailed is password protected.

Staff Induction when new staff are inducted they are asked to complete various forms containing personal details for their staff record and so that they can be paid. These forms may be scanned and emailed to finance so that staff can be set up on the system. These files may not always be password protected (e.g. if they are PDF scanned copies);

Staff leave (holidays) staff may request leave via email and advise colleagues of dates booked via email;

Office notice board staff on-duty/on leave may be noted on notice board in office. Although the office is locked, anyone who enters (clients or guests) will be able to see the notice board;

Staff Handover every day a handover note is created electronically for keeping Staff on the next shift (day or night care) informed of what has occurred during that day/night so they are aware of the history and can respond appropriately if something further occurs:

- This handover note is password protected and emailed and circulated to staff at the end of each shift;
- the handover contains client first names but not surnames, it may refer to third parties if they have been involved in an occurrence of relevance;
- the password is rarely changed (it was not clear from the stories collected whether there is a set procedure for changing the password);
- handovers are read through verbally at the end/beginning of each shift and both staff members from both the exiting and beginning shift sign the handover to indicate this has been done. This signed handover note is then stored in a folder within the office (locked office);
- there is also an electronic copy of the handovers stored in a shared folder on the internal drive, this can be accessed by all staff.

Email exchanges both internal and external. Email is used by most staff as part of their work to liaise with each other and external stakeholders about client care. The following issues/practices were noted:

- a number of these emails do contain personal details about clients, sometimes these details are contained in password protected attachments (e.g. staff handovers and client registers are password protected);
- alternatively, personal details may form part of the main body of the email, e.g. where exchange between external professional support worker and internal staff member or confirmation of funding arrangements for client from funding body;
- emails concerning client care are printed out and a hard copy placed in the client file;
- each staff member will manage their own email inbox, they are advised to ensure they delete any old emails.

Working from home Although the majority of data handling took place on StreetScene premises, some staff do work from home using home computers and/or mobile phones;

- Some staff accessed these emails from home computers and/or mobile phones;
- some staff work from home permanently and have dedicated computer used only for work which is password protected;
- some staff work from home occasionally, using their home computer or laptop. This may or may not be password protected;
- a number of staff check and respond to emails from home and or their personal phone, it is highly likely that data may be stored on home computer or personal phone as a result (e.g. auto download attachments).

D.5 Data Wheel - post-evaluation final questions

Q1 (Explanation, Description) What is the purpose of the data collection/processing?

Q2 What is the system, process, project or dataset about (subject matter/context)?

Q3 (Attributes) What is the lawful basis for processing the data?

Q4 If special category data processed, please also provide secondary lawful basis for processing the data

Q5 Please describe what data will/has be/been collected?

Q6 - 9 What attribute types will be included: PI, QI, Sensitive, non-sensitive? (attributes/dataset) - please select all that apply. If this is a new project or system, please also complete the Data Register worksheet with details of each attribute.

Q10 (Transmission Principles) How will the data be processed internally within the organisation?

Q11 Please describe the information flows with internal stakeholders (then please go to life of the form worksheet and complete this to gather more detailed information about the data collection/project/system being assessed

Q12 Will the data be transmitted to external stakeholders? (transmission principles)

Q13 Please describe the information flows with external stakeholders. Please confirm who is:

Q14 (Actors - internal) The data controller

Q15 Responsible Person: name and contact details of the designated Data Controller(s)

Q16 Data Owner (if different from Controller)

Q17 Data Subject(s)

Q18 Name & contact details of Data Protection Officer (DPO) if applicable

Q19 Other (please explain role and relationship)

Q20 (Actors - external) Please confirm who is the data controller

Q21 Please list any Data Processors who will be processing the data on behalf of the Data Controller (e.g. Cloud based software suppliers or other contracted suppliers who may hold or process the data for or on behalf of the Controller)

Q22 What is the relationship between the Data Controller and the Data Processor?

Q23 Is a contract in place between the Data Controller and the Data Processor?

Q24 Please list any joint Data Controller(s) (e.g. external partner organisation) who will be processing the data on behalf of or in place of the Data Controller

Q25 What is the relationship between the Joint Data Controller and the Data Controller?

Q26 Is a contract in place between the Joint Data Controller and the Data Processor?

Q27 Please list of all/any third parties with whom the data will be shared (these will be classed as external data processors) you may wish to add extra rows for each additional data processor(s)(External Data Processor(s))

Q28 External Responsible person, name and contact details of the designated External Data Controller

Q29 External Data Owner (if different from Processor)

Q30 External Data Subject(s)

Q31 Name & contact details of third-party Data Protection Officer (DPO) if applicable

Q32 External Other (please explain role and relationship)

Q33 (Prevailing Context) What are the relationships between actors? (please describe role and relationships between data- controller, processor(s) and Subjects)

Q34 What was/is the social context of the data collection? e.g. school would be educational context, council tax would be for tax collection etc.

Q35 Is/was data collected directly from data subject?

Q36 Has/will permission/consent been/be sought for processing of the data from data subject(s)?

Q37 Are there any overriding considerations as to why processing should be allowed despite lack of consent/limited consent?

Q38 Was/is the data collected with a view to process beyond its original purpose? (if yes, please answer questions below)

Q39 Does data originate from third party source? (if so, please state origin of the data)

Q40 Was consent sought from the data subject for secondary processing purposes? And were there any limitations on secondary purpose to which consent was given?

Q41 Has/will the data subject been/be informed of their rights to withdraw consent?

Q42 Are there any overriding considerations as to why processing should be allowed despite lack of consent/limited consent?

Q43 (Risk Assessment, What/Why) What are the benefits of the processing for the data subject?

Q44 What is/are the desired effects of the processing for the data subject?

Q45 What is the desired effect of the processing for the organisation?

Q46 What is the primary legal basis for collecting/processing or handling the data?

Q47 If applicable, what is the secondary legal basis for collecting/processing or handling the data?

Q48 What are/is the desired effects of the processing for the organisation?

Q49 Why is the data being collected/processed? (purpose limitation/relevance)

Q50 (How) How much data will be collected?

Q51 Who is responsible for security around manual data handling, processing or storing?

Q52 How will the data be collected?

Q53 How is/will the data be accessed and used/processed?

Q54 (Extent) What is the extent of the processing, will we require consent?

Q55 Identify any risks of the dataset or individual attributes, being obtained or accessed by unauthorised parties or means in such a way that they can pose a risk a data subject? (please be specific as to what risk and how this might pose a new risk)

Q56 Security - how is/will the data be protected and kept safe?

Q57 (Exposure) Who is responsible for access control and data security (electronic data handling, processing or storing)?

Q58 What are the risks of the dataset or attributes, being linked to external data in such a way that they can pose a risk of contributing to re-identification of a data subject? (please be specific as to what risk and how this might pose a new risk)

Q59 How will data handling boundaries be set, measured and controlled?

Q60 (Expectations How is/will compliance be measured and controlled?

Q61 Data subject right to access and erasure requests, how is/will this be managed/accommodated?

Q62 How is/will compliance be evidenced?

Q63 How will we meet stakeholders expectations and adhere to data subjects rights?

Q64 (Life Cycle) What is the expected data life cycle? i.e. for how long will the data be *live* and processed

Q65 Has data limitation been considered, is all the data being collected necessary?

Q66 How is/will the data be stored?

Q67 (Life of the form) What journey(s) will the data take as part of its lifecycle?

Q68 How do/will you delete the data?

Q69 How long will the data be kept?

Context (Surrounding Context) Could the data collection or processing be perceived to infringe upon:

Q70 any political values? (please also explain whose values and how they might be infringed upon)

Q71 any social values? (please explain what those values are, whose values they are and how these might be infringed upon)

Q72 any moral values? (please explain what those moral values are, whose moral values and how these might be infringed upon)

Q73 any legal compliance? (if yes, please clarify what legal compliance might be breached, how and who would be affected)

Q74 any belief systems that could adversely be affected by processing? (please explain what belief system could be affected, how and who would might be adversely affected)

Q75 What are the positive values that data processing will bring/enhance. These may include commercial gain, improved transparency, meeting legal obligation etc.

Q76 impose any form of power imbalance? (please explain how such a power balance might arise, who might be affected and how these might be imposed)

Q77 pose a threat to the autonomy or freedom of the data subject(s)? (please explain how this might pose a threat, who could be affected and how)

Q78 result in any form of discrimination (please explain what form of discrimination, how this might occur and who would might be adversely affected)

Q79 Grant or afford any privileges or prerogatives to the referer, the data- processor(s), controller(s) or originator(s) that may benefit or be perceived to benefit those parties as a result of processing, sharing or publishing the data?

Q80 Are there any overriding legal, moral or ethical reasons why processing should be allowed even if there is a risk of re-identification?

Q81 Is there a reasonable expectation on the part of the data subject(s) and/or data controller(s) of data being kept confidential and not shared?

D.6 Participant Information Sheet - Training and Workshop

The Participant information sheet used for the training sessions and the seminar sessions held as part of the one day workshop, states: *"Charities, like many organisations, struggle to understand how best to interpret and implement the new General Data Protection Regulations (GDPR) that came into force on 25th May 2018.*

Bournemouth University (BU) has been working with a local charity ("the Charity") assisting with putting in place suitable processes and practices to implement GDPR and demonstrate compliance. As part of this a Data Protection Impact Assessment (DPIA) Framework has been devised, the DPIA Data Wheel.

This DPIA Data Wheel was created so that the Charity is able to document and keep a record of the outcomes of each DPIAs they conduct. More importantly, the DPIA Data Wheel has been devised with any Charity in mind and can be used by any charitable organisation to conduct DPIAs and demonstrate compliance with GDPR going forward.

For example, for the Charity that took part in this study, the records they now have of the DPIAs conducted so far will aid not only in demonstrating compliance with GDPR to the Information Commissioners Office (ICO), but also the Care Quality Commission (CQC) who have also indicated that they must be kept informed as any breach could affect the care of the clients the charity work with.

Having created a DPIA process for assessing the privacy risks of any new and existing projects, we are going to go over this today. In particular, today's Seminar session is an opportunity for you to understand the DPIA Data Wheel and how to use this in conducting your own DPIAs going forward. To this end, you will be provided with a Scenario of a process the Charity use for gathering personal information about their clients.

The plan for today is:

1. Read the Scenario in the leaflet being handed out;

2. Discuss the responses to the DPIA Data Wheel for this Scenario. In particular, the questions pertaining to context;

3. Discuss what risks might be associated with the Scenario and complete the risk register for the Charity.

If you have any questions or queries, please feel free to either speak with your facilitator or contact the lead Researcher direct via email on ....

For research and ethical purposes, we also need for you to complete a participant sheet that explains how the data we collect will be used as part of our research. This form is at the end of this document. Please complete and sign this form and return it to us together with the completed user story questionnaire.

Finally, we would like to take this opportunity to thank you very much for assisting us with this work, we really value your input, time and comments."

D.7 Workshop Agenda - GDPR for Charities: How to conduct a DPIA

9:00 Coffee

9:15 Opening

9:30 Basics of GDPR & overview of the DPIA Data Wheel

10:00 DPIA worked example

10:15 Coffee Break

10:30 Seminar Groups: practical session conducting DPIA

12:00 DPIA review - seminar groups to report outcomes from sessions

12.30 Lunch

13:30 Charity representative gives talk about the experience of working with the research team

14:00 Finding security & privacy risks - talk

14:30 Panel discussion

15:30 Coffee Break

16:00 Next steps

16:30 Close

D.8 Facilitator Instructions

Facilitator handout - DPIA Seminar Sessions

You have kindly agreed to act a facilitator for one of the DPIA Seminar Sessions that are scheduled to be part of the GDPR for Charities: How to conduct a DPIA Workshop on Monday 11th June 2018.

First, thank you very much for agreeing to help out on the day.

We are expecting 50 participants on the day and are planning to have 4 seminar groups. Therefore, you will be facilitating a group of approximately 12 participants. The participants will all be staff or volunteers working in a not-for-profit organisation with an interest in understanding more about GDPR and how to conduct DPIAs.

These are your instructions for how to conduct the seminar, please do not hesitate to ask however, if you have any questions or need any clarification on any aspect of these instructions, the seminar or anything else.

Your role

Your main role as Facilitator is not to answer the questions, rather to facilitate and guide the conversation and perhaps give pointers so that the participants come up with the answers and suggestions.

You will however, be making notes of the responses in the spreadsheet that you will be provided with. These notes that you capture will be used as part of my research so, once again, I am most grateful for your help with this.

Handouts for Participants:

- Participant information sheet and participant agreement form.
- The information sheet gives participants brief details of the session and what we are trying to achieve and the participant agreement form is basically asking for their consent for me to use the outcomes of the seminar as part of my research. A copy of this form is attached to these instructions for your information
- Participant handout.
- Prior to the seminar session, participants will have been talked through GDPR and the DPIA Data Wheel. The handout is a reminder of these concepts and provides participant with information about GPDR and the DPIA Data Wheel. Again, a copy of this handout is attached for your information.
- Evaluation form. This is a basic questionnaire asking for feedback on the session, copy attached for your information.

Facilitating the session itself

The session itself is about the DPIA Data Wheel. The scenario participants have been given is one that seeks to take a predominantly paper driven process and making it electronic, using a tablet. Changing this process will change the risk landscape for what might or might not happen when the tablet is used for capturing the data and this is what the session is about.

The participants will have already been talked through the concepts and shown a demonstration of the DPIA Data Wheel.

At the start of the session

The first thing I would like you to do is to hand every participant a participant sheet and ask them to complete the participant form on the reverse of the participant sheet, sign it and return it to you before you start the session.

Once you have gathered the participant agreement forms, please give each participant a handout for their use during the session.

1. Explain to participants that the session will be in two parts, each consisting of about 30 minutes discussion around first, the DPIA Data Wheel Questions and second, the risks in changing the process.

2. Also explain that at the end of the session all the seminar groups will go back to the main seminar room for a re-cap session where each group will present a brief overview of the main risks they came up with in the session.
3. If possible, identify and appoint a note taker for the session who will note down the main risks agreed on in part two and a presenter who will present the main risks identified in the session to their peers in the recap session that will be held straight after the seminar. You may to leave this part until nearer the end. I will leave that up to you to decide which is going to work best for you.

Part 1 - Data Wheel Review

Please allow approximately 30 minutes for this part of the discussion - at the end of half an hour bring the discussion to a close and move to part 2. For this section use the Data Wheel worksheet on the spreadsheet.

1. To start with what I would like you to do is conduct a review of the original Data Wheel that was completed for the paper-based process. This involves talking through the Data Wheel answers that the Charity provided for the paper based existing process. These answers they gave have been included in the handout that you have given each participant. You will have a spread sheet based version of the same information and basically, it is a case of asking each question and then letting the group decide whether the change in the process will change the answer to that question and, if so, how.
2. Be aware that the handout does not include all the questions and answers - the ones not in the handout have been highlighted in green on your spreadsheet
3. Please note down any changed answers on your spreadsheet as the group works through the questions.
4. In particular, I would like you to discuss the questions around context both in the data and the wheel, this is where most of the debate is likely to take place as different people have different tolerances, values and norms.

Part 2 - Risk Register

Please allow approximately half an hour for this part of the discussion being mindful that the evaluation sheets will need to be handed out and completed by participants by 12. I would therefore suggest you wrap up this session at around 11.45 to allow time for nominating a speaker for the review session back in the main seminar room and the evaluation part. For this part of the seminar session use the Risk Register worksheet on the spread sheet

For the second part of the seminar we are concerned with ascertaining what the risks might be in changing the process from paper to electronic.

- Please start off the discussion with the group around what risks to privacy there might be in the new process. These by their nature will include security risks as any security breach could result in a privacy breach and vice versa.

- For each risk, please note down the risk and then ask the group to score the risk on severity, likelihood and impact for each score, there is a drop down list on the spreadsheet.
- For each risk, ask the group what they consider might be good mitigation strategy to eliminate or reduce the risk and note any responses on the spreadsheet.
- Towards the end of the session (around 11.45) wrap up the discussion and work with the group to agree which of the risks discussed are the main risks they have identified as part of the session and who will present these.

At the end of the session

Please give each participant an evaluation form and ask him or her to complete this and return it to you. Each session will be numbered so if they could indicate the session number on the form that would be helpful.

D.9 Training/Workshop Handout

D.9.1 GDPR Principles

Principle 1 Lawfulness, fairness and transparency:

- Lawful: Determine and define what is your lawful basis for processing the data;
- Fair: use & process the data fairly with data subjects interest in mind;
- Transparency: Inform the data subject(s) what data processing will be done.

Principle 2 Purpose Limitation:

- Data should be obtained only for specific, lawful purposes;
- Only collect relevant and necessary data, and process data fairly with data subjects interest in mind.

Principle 3 Data Minimisation:

- Data collected should be adequate, relevant and not excessive;
- This means you should collect minimum data i.e. only collect data that is needed for your specified purpose.

Principle 4 Accuracy:

- Keep data up to date and accurate

Principle 5 Storage Limitation:

- Anonymise or pseudonymise data as soon as practicable and keep for no longer than absolutely necessary;
- Securely delete and/or destroy data no longer needed.

Principle 6 Integrity and confidentiality

- Ensure appropriate security measures in place;
- Process and store the data securely ensuring data is protected from harm, unauthorised or unlawful access;
- Ensure all staff are trained in how to safely & securely handle and process data.

D.9.2 Data Protection Impact Assessment (DPIA): When should we conduct a DPIA

DPIA Compulsory

New Technologies Using new and/or innovative new processes or technologies;

Profiling/Automated Processing Use of extensive or systematic profiling or automated processing or decision making relating to the data subjects upon which significant decisions about people may be based;

Large Scale Sensitive Data Large scale processing of criminal offence or special category data;

Extensive Systematic Monitoring Large scale systematic monitoring of a public space or publicly accessible space;

Genetic/Biometric Processing of genetic or biometric data;

Data Linking Any form of linking, matching or combining of data from several sources;

Tracking Conducting processing or collect data that involves any form of tracking of behaviour and/or location of people either on- or off-line;

Children Providing online services to children and/or process data relating to children for marketing, profiling or automated decision-making

Potential Harm Conduct processing that involves personal data that, if a breach occurred, could result in mental, physical or other harm to the data subjects;

Evaluation/Scoring The processing of personal data to be used for scoring or evaluation.

DPIA Advisable

Across Borders Contract Offering goods or services to data subjects or monitoring their behaviour in multiple EU member states;

Free Movement across Borders Conduct any form of processing that could substantially affect the free movement of data subjects with the EU;

Systematic Monitoring Conducting any type of systematic monitoring;

Sensitive Data Processing of highly sensitive or personal data;

Large Scale Processing Any large-scale processing of data;

Vulnerable people Processing of any type of data relating to vulnerable people;

New Technologies Planning to use, implement or involve the use of emerging or innovative new technological or organisational solutions;

Withholding Access Any form of data handling that involves preventing the data subjects from gaining access to or using a service or contract or exercising a right.

D.9.3 The DPIA Data Wheel

	The Need for a DPIA	
D	**Data**	Will a new system, project or process be implemented that involves collecting, processing, transmitting, sharing and/or storing of data or are there any changes to an existing system, project or process?
P	**Protection**	Looking at the ”when should we conduct a DPIA” list opposite, will the system, project or process involve processing or using any of the types of data listed in the first column (”DPIA Compulsory”)
I	**Impact**	Looking at the ”when should we conduct a DPIA” list opposite, will the system, project or process involve processing or using any of the types of data listed in the second column (”DPIA Advisable”)
A	**Assessment**	Based on the answers above, DPIA required?

Table D.1: DPIA Data Wheel - pre-assessment

DATA			answers provided from Evaluation 2
D	**Description**	What is the purpose of the data collection/processing?	Rehabilitation; Planning the care to be given to meet clients needs
		What is the system, process, project or dataset about (subject matter/context)?	C - Combination of all (see below) i.e. all types, including personal data and highly sensitive data
A	**Attributes**	Please describe what data will/has be/been collected? (you may wish use/refer to the Data Register for more detailed data collection/guidance)	Used for counselling team to make decisions with the client about their care
T	**Transmission Principles**	How will the data be processed internally within the organisation?	Internally: Paper based
		How will the data be transmitted between stakeholders internally and externally? (you may wish use/refer to the life of the form for more detailed data collection/guidance)	Externally: fax, email, verbal, statistics (reporting)
A	**Actors**	Please list all staff/roles (actors) who will be processing/using the data	Internal: all Staff; External: Parts of the data will be shared with external stakeholders involved in caring for the client such as Care Manager, Doctor, Social Worker
		Please record who is the data processor(s), data controller, data subject and any other internal and external stakeholders who will use/access the data and what their role is	
Context	**Prevailing Context**	What are the relationships between actors? (please describe role and relationships between data- controller, processor(s) and Subjects)	Client/professional: no personal friendship or contact allowed for 2 year after treatment ceased. Do have staff working at the Charity that are former clients (more than 2 years ago); Professional/professional; Colleague/colleague may be friend/friend
		Was/is the data collected with a view to process beyond its original purpose?	No
		Does data originate from third party source (if so, please state origin of the data)	Sometimes, care managers send initial identification of needs. Courts, GP's, health professionals may send data to help us provide suitable and safe care
		Is/was data collected directly from data subject?	
		Has/will permission/consent been/be sought for processing of the data from data subject(s)?	Yes
		Was consent sought from the data subject for secondary processing purposes?	n/a
		If yes, were there any limitations on secondary purpose to which consent was given?	
		Has/will the data subject been/be informed of their rights to withdraw consent?	Yes
		Are there any overriding considerations as to why primary processing should be allowed despite lack of consent/limited consent?	Potentially if there are safeguarding issues relating to client or others at risk
		Are there any overriding considerations as to why secondary processing should be allowed despite lack of consent/limited consent?	n/a

Table D.2: DPIA Data Wheel - DATA

WHEEL			answers provided from Evaluation 2
W	**What**	What is the purpose of the data collection/processing?	to provide safe and effective rehabilitation from addiction
	Why	Why is the data being collected/processed?	to provide individual treatment programs
		What are the benefits of the processing for the data subject?	Recovery from addiction
		What are the benefits of the data processing for you (the organisation)?	Providing recovery treatment is what the organisation does
		What is/are the desired effects of the processing for the data subject?	For Client's to have a life in recovery
		?What is/are the desired effects of the processing for the organisation?	Quality of care, further funding for more treatment
H	**How**	How will the data be collected?	Paper
		How much data will be collected?	A lot
		How is/will the data be accessed and used/processed?	Paper, used for treatment only, e.g. data may be used by counsellors to ground decisions about care needed
		Who is responsible for security around manual data handling, processing or storing?	Staff
E	**Extent**	What is the extent of the processing - will we require consent?	Yes, granular consent sought
		Security: how is/will the data be protected and kept safe?	Locked away in filing cabinet in locked office
		How will data handling boundaries be set, measured and controlled?	
		Who is responsible for security around manual data handling, processing or storing	Staff
E	**Exposure**	Who is responsible for access control and data security (electronic data handling, processing or storing)? How is/will compliance be measured and controlled?	Managers checks, quarterly audits
E	**Expectations**	How is/will compliance be evidenced?	Managers checks, quarterly audits
		Data subject right to access and erasure requests, how is/will this be managed/accommodated?	Protocol and procedure being developed
		How will we meet stakeholders expectations and adhere to data subjects rights?	Protocol and procedure being developed
L	**Life Cycle**	How is/will the data be stored?	Paper format, stored in a locked cabinet in a locked office
		How much data will be collected?	Data only used for treatment, no other purposes.
		Life of the form - what journey(s) will the data take as part of its lifecycle?	Data only used for treatment, no other purposes. See transmission principles for how it may be shared
		Data storage and retention?	Data kept for 7-years, legal requirement to keep for this period
Context	**Surrounding Context**	Could the data collection or processing be perceived to infringe upon any social or moral roles, values or norms? (please explain whose values, roles or norms and how they might be infringed upon)?	Potential power imbalance in that staff have power to ask clients to leave if they break rules and endanger other clients; Some autonomy limitations are necessary in early stages of treatment e.g. during detox; Clients not allowed to go into places selling alcohol , safeguarding; Positive values of data processing would be abstinence, which would improve life quality for both the client and their social circle

Table D.3: DPIA Data Wheel - WHEEL

D.10 Training/Workshop Evaluation Form

All information provided will be used to inform and improve future training sessions. For each of the questions below, please circle the response that best describes how you feel about the statement:

Overall How do you feel your knowledge has improved regarding:

- GDPR? (Yes/ No)
- DPIAs? (Yes/ No)
- Would you be comfortable explaining GDPR to your friends and Colleagues? (Yes/ No)
- Would you be comfortable explaining what a DPIA is to your friends and Colleagues? (Yes/ No)

Rating today's session:

- Overall how would you rate today's training session? (Excellent / Very Good / Good / Fair / Poor)
- How much of the information provided today on GDPR was new to you? (All of it / Most of it /About half / Some of it / None of it)
- How likely are you to now use and apply the knowledge you have learned in your daily work (Extremely likely / Very likely / Moderately Likely / Slightly likely / Not at all likely)
- How much of the information provided to you on DPIAs was new to you? (All of it / Most of it /About half / Some of it / None of it)
- How confident would be in conducting a DPIA for a new or existing project? (Very confident / Mostly confident / Reasonably confident / A bit hesitant / Not confident at all)
- How likely would you be to recommend this training session to a colleague or friend? (Extremely likely / Very likely / Moderately Likely / Slightly likely / Not at all likely)

Please answer the following questions using your own words:

- What did you like about today's session?
- What do you feel went well?
- What did you dislike about today's session?
- Are there any aspects that you would like to learn more about or that you feel could be clarified better?
- Any other Comments/Suggestions?

D.11 Facilitator De-brief

Facilitator debrief - DPIA Seminar Sessions

Thank you very much for agreeing to and helping out on the day with the DPIA Seminar Sessions that took place on Monday 11th June 2018 as part of the GDPR for Charities: How to conduct a DPIA Workshop.

I hope you enjoyed the day, the session and the round the table discussion within your seminar group.

Please could you send me the completed spread sheet with the Data Wheel answers and risks identified within the group as part of the session.

I would also appreciate it, if you would take a few minutes to provide me with some feedback on what worked, what did not work and what you believe should be changed/added or removed for future sessions.

Facilitator Feedback:

Please answer the following questions about your experience as a facilitator on the DPIA Data Wheel:

- How comfortable were you that you had been given sufficient details and direction for facilitating the session?
- Were there any aspects of the instructions that you were provided with prior to the session that you feel were omitted/should have been included and, if so, please explain
- What do you feel went well?
- What do you feel did not go so well?
- What did you dislike about the session ?
- Are there any aspects of the session that you feel should have been omitted as part of the session?
- Any other Comments/Suggestions?

D.12 Staff Training Session Evaluation Results

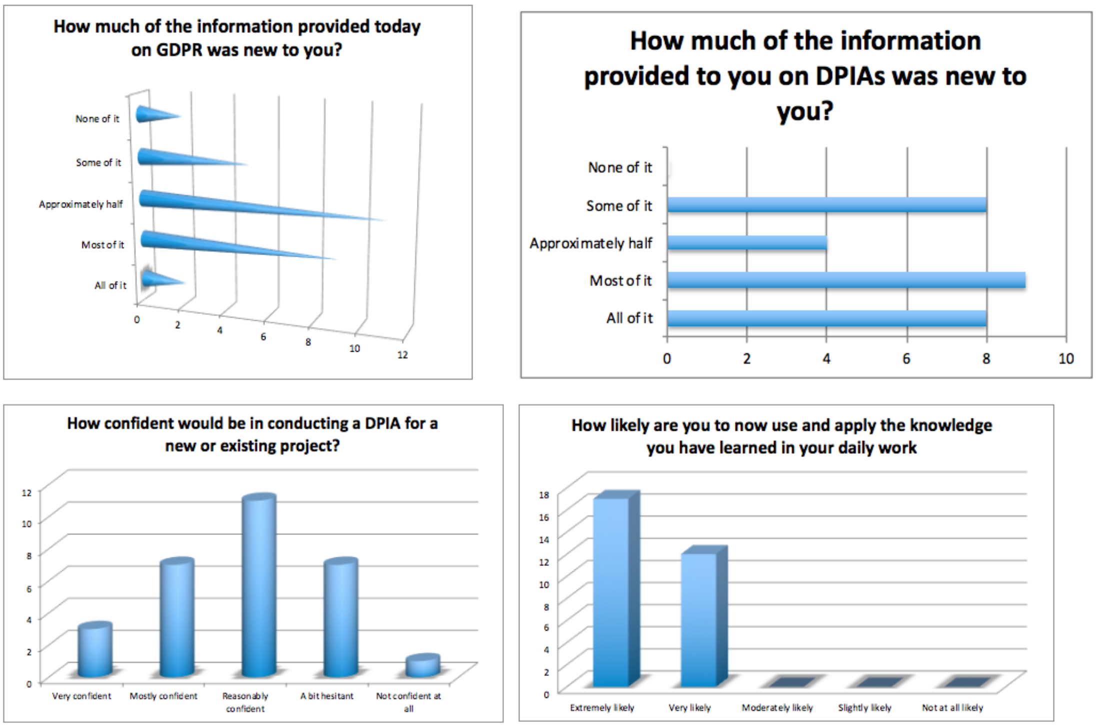


Figure D.5: Staff Training Evaluation Responses

D.13 Workshop - Evaluation Results

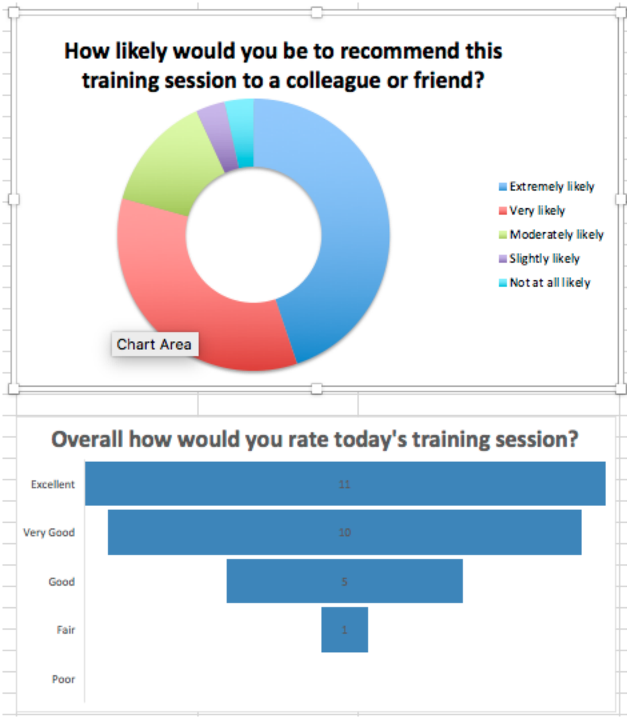


Figure D.6: Staff Training Evaluation Responses

	Overall How do you feel your knowledge has improved regarding GDPR	Overall How do you feel your knowledge has improved regarding DPIAs	Would you be comfortable explaining GDPR to friends & colleagues?	Would you be comfortable explaining what a DPIA is to friends & colleagues?
Yes	29	27	21	18
No	0	1	5	6
Other/Blank	0	1	3	5
Total	29	29	29	29

Table D.4: Staff post-training Evaluation Results

	Overall How do you feel your knowledge has improved regarding GDPR	Overall How do you feel your knowledge has improved regarding DPIAs	Would you be comfortable explaining GDPR to friends & colleagues?	Would you be comfortable explaining what a DPIA is to friends & colleagues?
Yes	17	19	17	15
No	2	0	3	4
Other/Blank	1	1	0	1
Total	20	20	20	20

Table D.5: Post-workshop Evaluation Results

D.14 DPIA Data Wheel - Instructions & Glossary

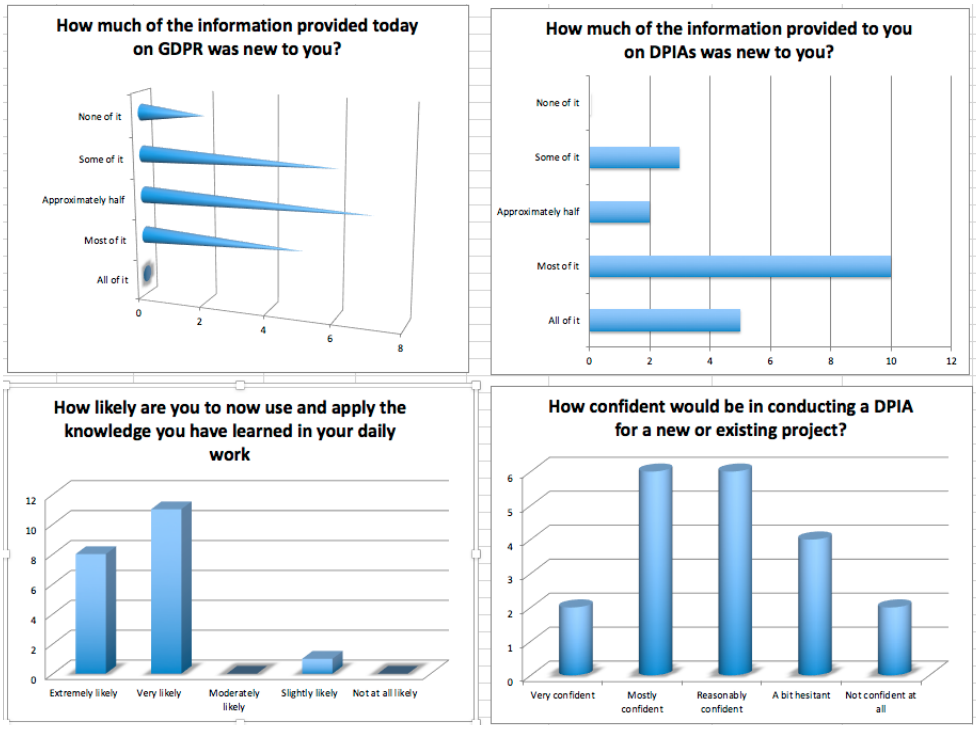


Figure D.7: Workshop Evaluation Responses

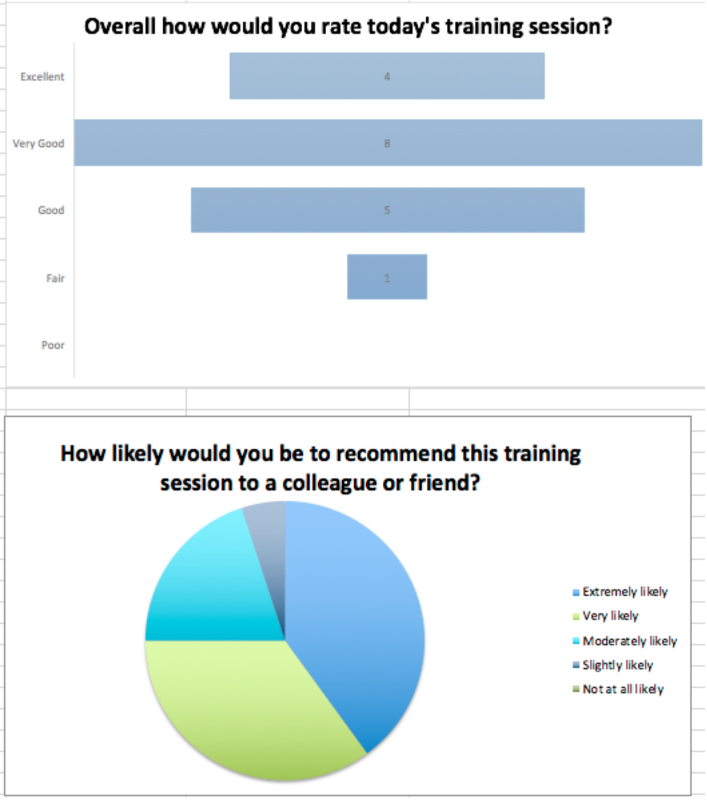


Figure D.8: Workshop Evaluation Responses

	Instructions
Worksheet 1- Need for a DPIA	This is the starting point of the DPIA framework. This is a short assessment that will help you determine whether or not a DPIA is required for your system, project or process. A YES answer here will mean conducting a DPIA and continuing through the rest of the framework. A NO answer, will end the process and ask you to provide an explanation and the reasons for why a DPIA is not considered necessary
Worksheet 2 - DPIA Data Wheel	Where a DPIA is required, the DPIA Data Wheel is the actual privacy risk assessment for the process, system or project. Here you will be asked to consider different aspects fo the process, system or project including what data you will capture, the people who will handle the data and how the data will be handled, updated, protected, shared and stored.
Worksheet 3 - Data Register	Here you will be asked to provide more specific and granular details about the data attributes (individual data items) that you plan to collect, process and handle. The information gathered here is designed to form part of your Master Data Register going forward, thereby helping you maintain an accurate overview of your organisation's data holdings. Therefore, the information gathered on this worksheet is intended to be used to compliment and help inform your risk assessment on Worksheet 2, the DPIA Data Wheel.
Worksheet 4 - Life of the Form	The questions on this worksheet have been designed to make you think about how the data travels within your organisation. By considering the "journey" the data within your system, project or process is likely to take during its lifetime, you will be able to glean valuable insight into where there may be potential risks that you will need to mitigate against. Therefore, the information gathered on this worksheet is intended to be used to inform your risk assessment on Worksheet 2, the DPIA Data Wheel.
Worksheet 5 - List	On this worksheet you can find a list of all the drop down menu's that form part of the assessment on the other worksheets.

Table D.6: DPIA Data Wheel Instructions

	Glossary	**please find below a glossary of terms, i.e. explanations for what some of the commonly used terms within the DPIA Framework: The DPIA Data Wheel mean**:
Term	*Meaning*	*GDPR Definition (Article 4)*
Data	**Terms Relating to Data**	
Personal Data	any information related to an individual (*identifiable natural person* i.e. the Data Subject), that may be used to directly or indirectly identify the person	Personal data means any information relating to an identified or identifiable natural person ("data subject")
Dataset ("Filing System")	the dataset, database or other data filing system where the data will be handled	means any structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis
Genetic Data	unique data relating the characteristics of an individual that is either inherited or acquired (e.g. DNA)	means personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question
Biometric Data	any personal data relating to the physical, physiological, or behavioural characteristics of an individual which allows their unique identification	means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data
Health Data	any data that relates to the health, mental or physical, of an individual. This includes providing health care services where personal data might be revealed	means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status
Data Handling	**Terms Relating to Handling the Data**	
Processing		means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction
Restriction of processing		means the marking of stored personal data with the aim of limiting their processing in the future
Profiling		means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements
Pseudonymisation		means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person

Table D.8: DPIA Data Wheel Glossary part 1

ACTORS	**Terms Relating to the people the data is about or who handle the data (the Actors)**	
Data Subject	the individual or person about whom the data pertains i.e. the person who the data being processed relates to/is about	
Individual /Person (”identifiable natural person”)	the data subject i.e. the person who the data being processed relates to/is about	an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location number, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.
Responsible Person (”Data Controller”)	the person responsible for making decisions about how the data is handled and who it is shared with	means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law
Data Processor	any person who handles the date i.e. collects, updates, maintains, stores, shares, view or accesses the data on behalf of the data controller	means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller
Stakeholder (”recipient”)	any person who is granted access to or with whom the data is shared	means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not
External Stakeholder (”third party”)	any external person who is granted access to or with whom the data is shared	means a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data
RULES	**Terms Relating to the rules around how data is handled**	
Consent of the Individual (”consent of the data subject”)	permission of the individual about whom the data pertains i.e. the person who the data being processed relates to/is about	of the data subject means any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her
Breach (”personal data breach”)	data has been accidentally or illegally obtained, disclosed or shared with a non-authorised person or persons	means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed.
Free Movement across Borders (”cross-border processing”)	data processing across multiple countries within the EU or conduct any form of processing that could substantially affect the free movement of data subjects with the EU	means either: processing of personal data which takes place in the context of the activities of establishments in more than one Member State of a controller or processor in the Union where the controller or processor is established in more than one Member State; or processing of personal data which takes place in the context of the activities of a single establishment of a controller or processor in the Union but which substantially affects or is likely to substantially affect data subjects in more than one Member State.

Table D.10: DPIA Data Wheel Glossary part 2

Appendix E

Supporting Documents - Ideal Cities Case Study: Facilitating PrivACy Throughout (PACT)

E.0.1 Ideal Cities Ethic's Application

Summary:

”This project is a case study designed to help organisations to informed decisions around privacy risks. The study builds on underpinning research from two previous case studies that devised and evaluate two privacy risk decision making frameworks based on contextual integrity:

a CLIFOD (ContextuaL Integrity for Open Data) a practical privacy decision framework for establishing whether data is suited for publication as open data. This framework guides helps users identify what privacy risks might be associated with publishing or sharing data with third parties. It was evaluated working with a UK local authority and proved effective. Further, it is contended this framework can be effective in assessing privacy risks for any data publishing or data sharing, not just for open data; and

b The DPIA Data Wheel, a data protection impact assessment framework devised to help organisations assess what privacy risks might be associated with processing of personal data and help them demonstrate compliance with the General Data Protection Regulation (GDPR) which came into force in May 2018.

In this case study, the above two frameworks will be amalgamated into one framework, a privacy tool framework, PrivACy Throughout (PACT), and an additional final element added, a privacy and lifecycle plan (PLAN) framework for the data.

BU has a working relationship with a number of organisations, municipalities and local authorities as part of the Ideal Cities project. The intention is that this case study will leverage off these contacts and seek to collaborate with one or more of these organisations, in particular, a communications company based in Cyprus, to trial the final PACT in practice.

This objectives of this project are:

1. Amalgamate and revise the two existing frameworks (CLIFOD and the DPIA Data Wheel) to create PrivACy Throughout (PACT), a privacy risk assessment framework that will support organisations for all their privacy risk decision making needs;

2. Evaluate the additional phase of the PACT, ”PLAN” to validate it’s application and effectiveness in practice.

The case study will be conducted by Jane Henriksen-Bulmer between October 1st and 31st. The PACT evaluation will entail applying the above privacy assessment approaches, developed by Henriksen-Bulmer, as part of her current doctoral research.

This project will transfer knowledge and tools to the collaborating organisations and the Ideal Cities project. This will help the organisations to make informed decision about their data and privacy risks, conduct privacy risk assessments for both existing and new projects, systems or processes in a consistent, repeatable manner and help facilitate demonstrable compliance with legal obligations under GDPR. Secondly, sharing the results of this project with Ideal Cities will help them develop links to a suitable a privacy risk assessment framework for the Ideal Cities platform they are developing. Further, a the intention is that Jane Henriksen-Bulmer and the Ideal Cities collaborators will write a journal article and/or conference paper based on the outcomes of this case study for publication as part of the Ideal Cities project.

The case study will utilise the existing participant form used in the previous two case studies copy attached) together with the attached participant information sheet.

Justification for collecting data

As part of the case study, only Meta data will be collected to identify the types and format of personal data the organisation uses and processes. However, no actual personal data will be collected beyond the participant forms that we ask staff to complete to agree to take part in the study, a copy of the participant form is attached. The only personal data that will be collected is the participant agreement forms which all participants will be asked to complete (copy attached). These forms will be retained by the researcher in paper format and kept in a locked cabinet. They will be retained until the end of June 2019 when it is anticipated that the corrected thesis will be deposited. After this date, all copies and papers will be shredded and securely destroyed. To ensure the project is a success it may become necessary to seek clarification on answers provided by participants. This will be done via email where possible but it may prove more effective in some instances to ask questions in person to gain clarification and in such instances, Interviews with stakeholders will be used. Any interviews will be taped and transcribed to ensure all information discussed is captured. To allow for this participants are asked to agree to this as part of completing the participant form. The participant forms and transcripts will be kept by the research team in a secure folder that only the research team will have access to. Any personal data captured on the participant forms will be anonymised prior to publication or sharing.”

E.0.2 Participant Information Sheet - Ideal Cities

Full title of the project: Ideal Cities Case Study: Privacy throughout the Data Lifecycle (PACT)

Invitation to take part: You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Who is organising/funding the research?: This work has received funding from the EU's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 778229 (Ideal-Cities).

What is the purpose of the project?: The General Data Protection Regulation (GDPR) has introduced mandatory requirements for organisations to assess privacy risk for any high risk processing of personal data. This includes any automated decision making, profiling, systematic monitoring or tracking activities that involves personal and/or sensitive data.

Why have I been chosen?: You have been chosen as an employee or associate of one of the Ideal Cities project partner organisations who have kindly agreed to help evaluate this framework and use this to document and keep a record of the outcomes of each privacy risk assessment they conduct. More importantly, the privacy risk framework has been devised to also help Ideal Cities partner organisations with planning and preparing for the data lifecycle once the data has been collected or processed, thereby helping to demonstrate compliance with GDPR going forward.

Do I have to take part?: It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a participant agreement form. You can withdraw from participation during the workshop or session at any time and without giving a reason. If you decide to withdraw we will usually remove any data collected about you from the study. Once the data collection activities have finished you can may still be able to withdraw your data up to the point where the data is analysed and incorporated into the research findings or outputs. At this point your data will usually become anonymous, so your identity cannot be determined, and it may not be possible to identify your data within the anonymous dataset. Withdrawing your data at this point may also adversely affect the validity and integrity of the research. Deciding to take part or not will not impact or adversely affect your treatment/care (or that of others).

What would taking part involve?: The idea is that we will work with you to complete:

1. A privacy assessment for an existing process, system or project to determine whether the data is suitable to be shared with third parties and, if not, whether any mitigation strategies can be implemented to make the data suitable for sharing (e.g. anonymising the data);
2. A DPIA for a new and/or changing data processing activity, system or project; and
3. Complete a Privacy Lifecycle PIAN (PLAN). This PLAN will use the information gathered and decisions made in as part of the above risk assessments to inform and help forward plan how the data will be managed, safeguarded and stored throughout the rest of its lifecycle with the organisation. This will involve establishing how the stages of the data lifecycle can be factored in as part of any forward planning for the data.

What are the advantages and possible disadvantages or risks of taking part?: Whilst there are no immediate benefits for those people participating in the project, it is hoped that the information gathered as part of these assessments will be used to help us better understand how data is collected, handled and shared both within the organisation, externally and with any third party stakeholders. More importantly, it will allow us to find out whether or not the privacy assessment framework is an effective tool for helping organisations conduct consistent, repeatable privacy risk assessments and plan appropriate privacy governance strategies for the data throughout its lifecycle with the organisation.

What type of information will be sought from me and why is the collection of this information relevant for achieving the research project's objectives?: The information gathered will be information about the data collected as part of the project or system and how this is processed and shared. This will be used to assess what privacy risks might be associated with the processing, handling and/or sharing of that data.

Will I be recorded, and how will the recorded media be used?: The audio recordings of your activities made during this research will be used only for analysis and the transcription of the recording(s) for illustration in conference presentations and lectures. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings.

If you have consented to your photograph being taken during sessions or workshops these may be included in research outputs or displays.

How will my information be kept?: All the information we collect about you during the course of the research will be kept strictly in accordance with current data protection legislation. Research is a task that we perform in the public interest, as part of our core function as a university. Bournemouth University (BU) is a Data Controller of your information which means that we are responsible for looking after your information and using it appropriately. BU's Research Privacy Participant Privacy Notice sets out more information about how we fulfil our responsibilities as a data controller and about your rights as an individual under the data protection legislation. We ask you to read this Notice (BU Research Participant Privacy Notice (https://www1.bournemouth.ac.uk/about/governance/access-information/data-protection-privacy)) so that you can fully understand the basis on which we will process your information.

Publication: You will not be able to be identified in any external reports or publications about the research without your specific consent*. Otherwise your information will only be included in these materials in an anonymous form, i.e. you will not be identifiable.

Security and access controls: BU will hold the information we collect about you in hard copy in a secure location and on a BU password protected secure network where held electronically. Except where it has been anonymised your personal information will be accessed and used only by appropriate, authorised individuals and when this is necessary for the purposes of the research or another purpose identified in the Privacy Notice. This may include giving access to BU staff or others responsible for monitoring and/or audit of the study, who need to ensure that the research is complying with applicable regulations.

Sharing and further use of your personal information: As well as BU staff [and the BU student(s)] working on the research project, we may also need to share personal information in non-anonymised for with [insert details or any third parties who may need to access the data and why e.g. external organisation(s) such as external collaborators, transcription services and funders.

The information collected about you may be used in an anonymous form to support other research projects in the future and access to it in this form will not be restricted. It will not be possible for you to be identified from this data. Anonymised data may be added to BU's Data Repository (a central location where data is stored) and which will be publicly available.

Retention of your data: All personal data collected for the purposes of this study will be held for maximum 1 year from the after the award of the degree. Although published research outputs are anonymised, we need to retain underlying data collected for the study in a non-anonymised form for a certain period to enable the research to be audited and/or to enable the research findings to be verified.

Contact for further information: If you have any questions or queries, please feel free to either speak with your supervisor or contact Jane Henriksen-Bulmer direct via email on jhenriksen-bulmer@bournemouth.ac.uk

In case of complaints: Any concerns about the study should be directed to Jane Henriksen-Bulmer. If you concerns have not been answered by Jane, you should contact Dr Shamal Faily, Senior Lecturer in Systems Security Engineering at Bournemouth University by email to researchgovernance@bournemouth.ac.uk.

Finally: If you decide to take part, you will be given a copy of the information sheet and a signed participant agreement form to keep.

Thank you for considering taking part in this research project.

E.0.3 Participant Agreement Form - Ideal Cities

Ideal Cities Case Study: Privacy throughout the Data Lifecycle (PACT)

Full title of the project

Name, position and contact details of researcher

Name, position and contact details of supervisor (if the researcher is a student)

PART A

"In this Form we ask you to confirm whether you agree to take part in the Project. We also ask you to agree to some specific uses of your identifiable information, which we will only do with your consent. You should only agree to take part in the Project if you understand what this will mean for you. If you complete the rest of this Form, you will be confirming to us that:

- You have read and understood the Project Participant Information Sheet (IdealCitiesV1) and have been given access the BU Research Participant Privacy Notice (https://www1.bournemouth.ac.uk/about/governance/access-information/data-protection-privacy);
- You have had the opportunity to ask questions;
- You understand that:
  - Taking part in the research might include being photographed; and/or being recorded (audio) on the basis that these recordings will be deleted once transcribed;

- Your participation is voluntary. You can stop participating in research activities at any time without giving a reason, and you are free to decline to answer any particular question(s);
- If you withdraw from participating in the Project, you may not always be able to withdraw all of your data from further use within the Project, particularly once we have anonymised your data and we can no longer identify you;
- Data you provide may be included in an anonymised form within an dataset to be archived at BU's Online Research Data Repository;
- Data you provide may be used in an anonymised form by the research team to support other research projects in the future, including future publications, reports or presentations.

Consent to take part in the Project Yes / No

I agree to take part in the Project on the basis set out above Yes / No (tickbox)

PART B

Consent to participating in specific Project activities Yes / No

I agree to be photographed during the Project Yes / No (tickbox)

Consent to use of information in Project outputs Yes / No

I understand that my words may be quoted in publications, reports, web pages and other research outpu Yes / No (tickbox)

Please choose one of the following two options:

I would like my real name used in the above Yes / No (tickbox)

I would not like my real name to be used in the above Yes / No (tickbox)

I agree for my photograph to be included in research outputs Yes / No (tickbox)

PART C

Name of Participant

Date & Signature

Name of Researcher

Date & Signature

Appendix F

PACT Layout and Questions

F.1 PACT Instructions & Glossary

The first worksheet in the spreadsheet contains the instructions for the framework. This is headed by a copyright notice that reads *Copyright (C) 2018 Jane Henriksen-Bulmer All title, including but not limited to Intellectual Property rights and copyrights, in and to the PrivACy Throughout (PACT) Framework and any copies thereof are owned by Jane Henriksen-Bulmer*

The following instructions then follow:

Step 1 - Overview This is the starting point for achieving PrivACy Throughout (PACT). This section asks you to provide an overview of the activity/scenario/project/system being assessed;

Data Register Here you will be asked to provide more specific and granular details about the data attributes (individual data items) that you plan to collect, process and handle. The register consists of three sections, A and B and C, to reflect that each assessor may be assessing data for different purposes e.g. they may not process any personal and/or sensitive data and therefore, you will have the choice to only complete section A (and possibly C) if that is the case;

Data Journey The questions on this worksheet have been designed to make you think about how the data travels. By considering the 'journey' the data within your system, project or process is likely to take during it's lifetime, you will be able to glean valuable insight into where there may be potential risks that you will need to mitigate against. Therefore, the information gathered on this worksheet is intended to be used to inform the rest of the framework;

Step 2 - Need for a DPIA This is a short assessment that will help you determine whether or not a DPIA is required for your system, project or process.
A YES answer here will mean completing an assessments of risks from the perspective of the individual (data subject) whose data you are handling (this assessment is in Step 4).
A NO answer, will mean you may skip Step 4 in which case you are asked to provide an explanation and the reasons for why a DPIA is not considered necessary.;

Step 3 - Wider Context This section begins the Risk Assessment. These questions relate to the surrounding context of the activity/scenario/project/system being assessed;

Step 4 - Risks to the Individual Here you will identify, note and score all the privacy risks from the perspective of the individual (the *Data Subject*. The risks identified should be derived from and informed by the information you have captured in the previous steps;

Step 5 - Risks to the Organisation This section forms part of the risk assessment. Here are asked to note and score all the privacy risks from the perspective of the organisaiton. These need to be considered in light of the information captured in the previous steps and any risks identified to the individual in Step 4;

Step 6 - PLAN This section has been included to help you plan how the data will be managed during its life time with the organisation. The data lifecyle is designed to encourage you to consider privacy throughout the data management life cycle. The data lifecyle has been used to delineate the different stages of data: collection, transformation, retention, access/release, post-access, disposal and governance and consultation. Further some guidance and ideas for how you may conduct consultation has been provided below;

List On this worksheet you can find a list of all the drop down menu's that form part of the assessment on the other worksheets;

CONSULTATION IDEAS Once you have completed the worksheets, it is important to consult on the risks with all your stakeholders. This can take the form of sending them the spread sheet for comments but it is suggested that the most effective way to consult is to hold meetings, workshops or seminars with groups of stakeholders to discuss the privacy risks associated with a particular system, project or process. You could for example, use this as an opportunity to:

- Raise staff / stakeholder awareness of privacy risk;
- Train staff/stakeholders on GDPR and privacy;
- Establish other risks not already identified that you were not aware of;
- Get feedback on existing processes and protocols within the organisation, what works and what does not;
- Find suitable mitigation strategies for dealing with the risks identified.

Staff/stakeholders who have been involved in coming up with the solutions are more likely to adhere to any new procedures and implement any changes identified to mitigate against the risks if they have been involved in devising these in the first place.

F.1.1 Glossary

Below the instructions is a glossary of terms used in the framework. This is more or less the same glossary used in the DDW with a few terms added. All the terms however, are also explained within the various steps where they relate. However, for completeness these are provided in Table F.1.

Term	Meaning	GDPR Definition
DATA	**Terms Relating to the Data**	
Attribute (Data Asset)	individual elements or classes of data within a dataset (e.g. name, date of birth, street etc.)	

Continued on next page

PI or ID Personal Identifiers or Identifiers	Personal Identifiers, data that can directly identify an individual such as their name; Use PI/ID where ANY individual affected, directly or indirectly, can be identified from the data item - this would include identifying a family member of the data subject	
Personal Data	Any information related to an individual (”identifiable natural person” i.e. the data subject) that may be used to directly or indirectly identify the person	Personal data means any information relating to an identified or identifiable natural person (data subject)
Dataset (”Filing System”)	A collection of data held in a data base or similar repository i.e. the dataset, database or other data filing system where the data will be stored or handled	Means any structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis
Genetic Data	unique data relating the characteristics of an Individual that is either inherited or acquired (e.g. DNA)	Means personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question
Biometric Data	any personally identifying behaviours or characteristics such as fingerprints or features that may be used to uniquely identify an individual	Means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data.
DATA HANDLING	**Terms Relating to the people the data is about or who handle the data (the Actors)**	
Annonymisation	Obfuscating or masking the data such that the data subject is no longer identifiable from the data	Means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person
Confidentiality	Ensuring data is only accessible to authorised stakeholders	
Data Availability	Ensuring data is usable on demand and accessible to authorised stakeholders	
Data Quality	Ensuring the data is updated, corrected as required and accurate	Every reasonable step should be taken to ensure that personal data which are inaccurate are rectified or deleted.
Dataset (”Filing System”)	see data section above	
Intervenability	Enabling data subject access and/or supervisory authority access to affect action on the records (e.g. request modification and/or deletion). In that way it can be seen as a safeguarding measure that must be included within any process or system involving personal data	
Processing the data	Means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.	
Profiling	Analysing data to identify or predict trends such as personal preferences, behaviour, location, movement etc.	Means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural persons performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements
Pseudonymisation	Replacing identifying data with pseudonyms ensuring any links to original data cannot be made by unauthorised parties	Means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person
Restriction of processing	Limiting future processing of the data by marking the stored data	Means the marking of stored personal data with the aim of limiting their processing in the future.
Unlinkability	Ensuring data cannot be sufficiently distinguished to linked across platforms or domains with similar context or purposes	
Unobservability /Undetectability	Ensuring data is anonymised so that a anonymity and undectability of the individual is preserved	
ACTORS	**Terms Relating to the people the data is about or who handle the data (the Actors or stakeholders)**	
Data Subject(s)	The individual(s) or person(s) about whom the data pertains - i.e. the person who the data being processed relates to/is about	An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person

Continued on next page

Individual /Person ("identifiable natural person")	The individual(s) or person(s) about whom the data pertains - i.e. the person who the data being processed relates to/is about	An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person
Data Controller	Means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law	
Data Processor(s)	Any third party entity or person who handles the date on behalf of the data controller i.e. collects, updates, maintains, stores, shares, view or accesses the data. Any processor must be contracted to conduct the work on behalf of the controller (Art 28)	Means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller
Joint Data Controller	A partner organisation who jointly manage the data with the Data Controller.	Where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers.
Responsible Person	A designated contact person with the organisation who acts on behalf of the Data Controller or Data Processor.	Means a natural or legal person established in the Union who, designated by the controller or processor in writing pursuant to Article 27, represents the controller or processor with regard to their respective obligations under this Regulation
Stakeholder ("recipient")	Any person who is granted access to or with whom the data is shared	Means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not.
External Stakeholder ("third party")	Any external person who is granted access to or with whom the data is shared	Means a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data.
RULES	**Terms Relating to the rules around how data is handled**	
Consent of the Individual ("consent of the data subject")	Permission of the individual about whom the data pertains i.e. the person who the data being processed relates to/is about	Of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.
Breach ("personal data breach")	Data has been accidentally or illegally obtained, disclosed or shared with a non-authorised person or persons	Means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed
Free Movement across Borders ("cross-border processing")	Data processing across multiple countries within the EU or conduct any form of processing that could substantially affect the free movement of data subjects with the EU	Means either: (i) processing of personal data which takes place in the context of the activities of establishments in more than one Member State of a controller or processor in the Union where the controller or processor is established in more than one Member State; or (ii) processing of personal data which takes place in the context of the activities of a single establishment of a controller or processor in the Union but which substantially affects or is likely to substantially affect data subjects in more than one Member State.
Proportionality assessment.	Proportionality requires that any limitation on the rights of the individual have to be justified. For example, making sure that the measure(s) taken in processing the data do not disproportionally limit the rights of the individual whose data is being processed. A pre-condition is that the measure(s) taken in processing or safeguarding are sufficient to achieve the objective while only relevant personal data for the purposes of the processing is collected and processed.	An assessment of the necessity and proportionality of the processing operations in relation to the purposes.
CONTEXT	**Terms Relating to context)**	
Norms	Norms are customs or expected rules, expectations or standards for how we are supposed to behave in society. Norms can relate to behaviours, ideas and beliefs etc. that have become embedded within us as individuals and society, i.e. influences and standards of expected behaviours or customs. Norms affect how we react, behave and perceive the world around us. Examples of norms include political or religious beliefs, an expectation of non-discrimination, equality (e.g. equal opportunity etc.)	
Prevailing Context	Prevailing context refers to the existing context, for example, if an existing dataset is being assessed, what is the current context within which data is processed? For this both the local context (e.g. within the organisation, say a telecommunications company) and the surrounding context (e.g. the national telecommunications network).	

Continued on next page

Values	Values are the underlying standards and specific guidelines that we apply to the norms. They are standards or principles of behaviour i.e. our judgement of what is important. Our values are what tells us what is good or bad, what is acceptable and what is not. Values include democracy, freedom of speech and autonomy. In terns of data, our values, societal and individual, will influence how we approach and handle the data which may, for example, be in the best interest for profit but not for the individual whose data is being processed.	
Wider Context	This refers to the surrounding context, how our norms and values help shape our actions and what effect this might create for the data that we are processing or handling.	

Table F.1: PACT Glossary of Terms

F.2 PACT Step 1 - Overview Questions & Guidance

The section is introduced with the following instructions: *This section asks you to provide an overview of the activity/scenario/project/system being assessed. This involves providing details of the processing of the data associated with this activity/scenario/project/system. Once complete please move to Step 2, to determine whether a Data Protection Impact Assessment needs to be conducted.* .

The questionnaire then commences asking 50 questions. These can be found in Chapter 8, Section 8.4.3). However, for completeness, these have also been repeated below:

Q1 Describe the data/dataset; What is the system, process, project or dataset about (subject matter/context)?

Q2 What is the purpose of the data collection/processing?

Q3 What is the lawful basis for collecting/processing the data (drop-down list). see Chapter 2, Section 2.4.4 for a list of these)

Guidance Q3 GDPR places a requirement on organisations to specify a lawful basis for processing of personal data.

Q4 If special category data processed, please also provide secondary lawful basis for processing the data (drop-down list). see Chapter 2, Section 2.4.4 for a list of these) Where special category data (e.g. data relating to health) is processed, a secondary lawful basis for processing may also be selected e.g. the primary lawful basis may be contractual with consent as a secondary lawful basis for certain elements.

Guidance Q1-4 This section sets the scene for the assessment and provides an overview of what data/system/process or project being assessed. *Processing* refers to ”any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction”.

Q5 Attributes - at a high level, please describe what data will/has be/been collected?

Q6 Please go to the data register worksheet and complete the data register with more detailed information about the attributes that will be collected/processed. If no personal or sensitive data will be collected/processed or stored, you may decide to complete section A only (however, completing section B is recommended for all data).

Guidance Q5-6 In order to understand the information flows it is necessary to first understand the data. What is the data about and what are the individual parts that make up the datasets. *Attributes* are the individual data elements within a data set. (see Section F.2.1 for column headers in the Data Register)

Q7 If the data/system/process/project being assessed is a a new practice please describe the data flows and how it is proposed the data will flow between actors? (i.e. any stakeholders using or processing the data)

Q8 If the data/system/process/project being assessed is a change to existing practice please describe the proposed changes in data flows and what changes in flow this will bring about?

Q9 If the data/system/process/project being assessed is a new practice please describe the data flows and how it is proposed the data will flow between actors? (i.e. any stakeholders using or processing the data)

Guidance Q7-9 This section seeks to find out about data flows, i.e. how the data is processed. Please answer the question that relates to the scenario being assessed i.e. Q7 for existing; Q8 for a change to existing process/practice; or Q9 for a new process/practice.

Q10 What is the overriding context within which this processing/collection of data takes place? e.g. public service, national telecoms, health care etc.

Q11 what is the sector context nested within the overriding context where this data collection practice takes place? e.g. library, telecommunications, teaching hospital.

Q12 What is the local (or departmental) context within which the data is processed/collected? e.g. provision of bin collection services, order processing, marketing, patient records system etc.

Guidance Q10-12 This relates to familiar level of context and the embedded context within that overriding context.

Q13 How is/will the data be processed internally within the organisation?

Q14 Please describe the information flows with internal stakeholders

Q15 Will the data be transmitted to external stakeholders?

Q16 If yes, please describe the information flows with external stakeholders.

Q17 Please go data journey worksheet and complete this to gather more detailed information about the data lifecycle of the collection/project/system being assessed (see Section F.2.2 for questions in the Data Journey).

Guidance Q13-17 These questions relate to how the existing information flows, i.e. how the data is currently processed.

Q18 Who is the designated data controller for the data? (normally this will be the organisation).

Guidance Q18 GDPR requires that organisations keep a record of who is DC.

Q19 Responsible Person - name and contact details of the designated person acting on behalf of the Data Controller(s).

Guidance Q19 This will be the contact person who acts on behalf of the DC.

Q20 Who is the Data Owner? (if different from the Data Controller).

Guidance Q20 In some instances, your organisation may may process data belonging to another entity.

Q21 Name & contact details of Data Protection Officer (DPO) (if applicable).

Q22 Who is the Data Subject(s)?

Q23 What is the relationship between the Data Subject and the Data Controller/Processor/DPO? (e.g. friend, co-worker, professionally, citizen, employment) Please note all relationships that apply.

Q24 Are any internal stakeholders also Data Subject(s)? (e.g. staff).

Q25 If there are internal data subjects, please describe the relationship between them and the data owner/controller and/or DPO?

Q26 How does the Responsible person(s)/Data Owner and/or DPO interact with the internal data subject(s)? (e.g. friend, co-worker, professionally, citizen, employment) Please note all relationships that apply.

Guidance Q18-26 Please note who the actors are. These are the stakeholders. These include the data controller (DC), Data Owner (DO), Data Protection Officer (DPO), Data Processor (DP), Joint Data Controller (JC) and the data subject(s) (i.e. the individuals whose data is being processed). Further details and descriptions are available in the instructions worksheet (under actors).

Q27 Please list any Data Processors who will be processing the data on behalf of the Data Controller (e.g. Cloud based software suppliers or other contracted suppliers who may hold or process the data for or on behalf of the Controller).

Q28 What is the relationship between the Data Controller and the Data Processor? (e.g. professional/friends/employer/employee etc.)

Q29 Is a contract in place between the Data Controller and the Data Processor?

Guidance Q27-29 Data Processors (DP) please note details of any Data Processor(s) or Joint-Data Controller(s) who will process/use/access the data for or on behalf of the Data Controller and what their relationship is with the organisation. This is relevant where you use the services of an outside organisation to process data on your behalf.

Q30 Please list any Joint Data Controller(s) (e.g. external partner organisation) who will be processing the data on behalf of or in place of the Data Controller.

Q31 Is a contract in place between the Joint Data Controller and the Data Processor?

Q32 What is the relationship between the Joint Data Controller and the Data Controller? (e.g. professional/friends/employer/employee etc.)

Guidance Q30-32 Joint Data Controller(s) (JC) This is relevant where you process data jointly with another entity.

Q33 External Responsible person - name and contact details of the designated External Data Controller/Processor.

Q34 Name & contact details of third-party Data Protection Officer (DPO) if applicable.

Q35 How does the Data Processor and/or Joint Data Controller interact with the data subject(s)? (e.g. friend, co-worker, professionally, citizen, employment) Please note all relationships that apply.

Guidance Q35 DP/JC Relationship to Data Subject.

Q36 Other Actors not mentioned (please explain role(s) and relationship).

Guidance Q36 Please note any other internal or external stakeholders who will use/access the data and what their role is.

Guidance Q33-35 Please note the contact person(s) within the external stakeholders who is responsible for the data use/access and what their role is within the organisation.

Q37 What was/is the social context of the data collection e.g. hospital would be health context, school would be educational context etc.

Q38 Is/was data collected directly from data subject?

Q39 If not, please note from whom the data was obtained and whether the data subject has been informed?

Guidance Q37-39 What are the existing informational norms with regards to the data, i.e. what is the prevailing social context.

Q40 Has/will permission/consent been/be sought for processing of the data from data subject(s)?

Q41 If consent has not been obtained are there any overriding considerations as to why processing should be allowed despite lack of consent/limited consent?

Q42 Was/is the data collected with a view to process beyond its original purpose? (if yes, please answer questions below)

Q43 Does data originate from third party source (if so, please state origin of the data)

Q44 Was consent sought from the data subject for secondary processing purposes?

Q45 Were there any limitations on secondary processing purpose(s) to which consent was given?

Q46 If no consent obtained for secondary processing, are there any overriding considerations as to why secondary processing should be allowed despite lack of consent/limited consent?

Q47 Has/will the data subject been/be informed of their rights to withdraw consent?

Guidance Q40-47 *Prevailing context* refers to the existing context, for example, if an existing dataset is being assessed, what is the current context within which data is processed? For this both the local context (e.g. within the organisation, say a telecommunications company) and the surrounding context (e.g. the national telecommunications network). Locate applicable entrenched informational norms and identify significant points of departure. Consider legal and ethical obligations e.g. consent, is it required/will it be obtained?

Guidance Q37-47 This section seeks to establish the existing context within which data is collected, processed, stored or shared.

Q48 Based on the contexts identified above, are there any potential impacts from how the data flows that could result in a violation of the privacy of the *data subject*?

Guidance Q48 This question seeks to establish whether *prima facie* the existing processes and data flow could have potential to violate the privacy of the data subject.

Q49 If the data/system/process/project being assessed is a change to existing practice (e.g. contracted third party processing on your behalf) please describe the proposed changes in data flows this change will bring about.

Q50 Based on the contexts identified above, *prima facie* is there any potential impacts from the proposed change in data flows that could result in a violation of *privacy of the data subject* or *compromise the confidentiality of the organisation*?

Guidance Q49-50 These question seek to establish whether, in light of any third-party data processing (DP contractors and/or JC's create a change in existing processes and data flows that could have potential to violate privacy.

F.2.1 PACT Data Register Questions & Guidance

Question 6 in the Overview section asks practitioners to complete the data register. This register will therefore become the third worksheet in the spreadsheet, named ”Data Register”. The columns and data to be captured here have been devised based on the DDW Data Register with some additional columns to account for both existing and new data being captured in the register. Further the order of the columns have been changed into two sections, A and B and C, to reflect that each practitioner may be assessing data for different purposes e.g. they may not process any personal and/or sensitive data and therefore, he/she will have the choice to only complete part A if that is the case.

The introduction section to this worksheet reads: *This section compliments the overview and provides the space for you to answer question 6. The information gathered here is designed to form part of your Master Data Register going forward, thereby helping you maintain an accurate overview of your organisation's data holdings. Therefore, the information gathered on this worksheet is intended to be used to compliment and help inform the rest of the framework. Here you are asked to provide more specific and granular details about the data attributes (individual data items) that you plan to process. The register consists of three sections, A and B and C, to reflect that each assessor may be assessing data for different purposes e.g. they may not process any personal and/or sensitive data and therefore, you will have the choice to not to only complete part A (and* *possibly C) if that is the case. However, it is recommended that you consider all the sections and justify your decisions as you work through the register.*

The following column headers are then provided below the introduction:

Section A This section must be completed for all assessments;

Attribute ID This is to allow each attribute to be given a unique ID, the format has been left open as this may be a number or text or a combination, depending on organisational preferences;

Attribute Name This is the individual data element, the name of the asset or data asset;

Optional: Attribute Description This is an optional column where practitioners can note details about the form/record/system data is collected for/on;

Attribute (Asset) Category This contains a drop-down menu where practitioners can choose between: (i) Personal Identifier (PI), (ii) Quasi-identifier (QI), (iii) Sensitive (s), (iv) Non-sensitive (NS) or (v) Other (O) if further investigation or second opinion required on the category for that attribute;

Data Type Here practitioners can capture the type of data and/or media type e.g. text, video, image etc.;

Data Collection Method This is to capture how the data will be collected e.g. paper, electronic system, web form etc.;

Storage Method Here details of how the data will be stored can be recorded;

Storage Time Period Record of how long the data will be kept;

Time Justification Asking for an explanation for why the data is stored for specified time e.g. legal obligation;

Destruction Method At the end of storage or processing how will data be destroyed;

Transmission Methods Please specify how data will be shared/transmitted;

Safeguarding Method provide a brief description for how the data will be safeguarded (you may wish to complete this after the assessment(s) are complete).

Section B Must be completed if any of the data collected contains personal and/or sensitive data that is being/will be collected/processed/stored;

Lawful Basis for processing Here a drop down menu allows practitioners to select which GDPR compliant legal basis for processing applies to the data;

Data Subject Categories Listing what categories of data subjects data will be collected from (e.g. employees, customers, clients, members etc.)

Purpose What is the purpose of the processing (this may be copied from the Overview);

Justification for Collecting/processing Here three columns have been allocated to this as there may be more than one reason for processing;

Responsible Person This refers to who within the organisation is designated 'Responsible person" on behalf of the Data Controller (DC) (this may be copied from the Overview);

DPO Contact details for who the DPO is (if applicable) (this may be copied from the Overview);

Recipient Categories This refers to any other third party recipients of the data (e.g. suppliers, customers, government agencies, credit referencing agencies etc.).

Section C - Complete this section if any third parties will be responsible for processing the data either on behalf of or jointly with the Organisation;

Data Processor(s)(DP) Note all data processor's who will process the data on behalf of the DC (this may be copied from the Overview);

DP Responsible Person To capture the contact details of the responsible person within any DP organisations (this may be copied from the Overview);

Joint Controller (JC) To capture details of any JC organisations (this may be copied from the Overview);

JC Responsible Person To capture the contact details of the responsible person within any JC organisations (this may be copied from the Overview);

Other Recipients This refers to any other third party recipients of the data not captured elsewhere (e.g. third party organisations outside the EU);

Other Recipients Responsible Person please provide details for the responsible person from any other recipients listed.

F.2.2 PACT Data Journey Questions & Guidance

Question 17 asks practitioners to complete the data journey. This is situated in the fourth worksheet within the PACT spreadsheet. It has been copied from the DDW's "life of the form" worksheet and renamed "Data Journey" to reflect that the data journey may not always refer to a form. Moreover, some of the questions asked have been revised slightly to reflect the change in name.

This worksheet is introduced with the following header: *This section compliments the overview and provides the space for you to answer question 17. These questions are intended to inform the rest of the framework. make you think about how the data travels. By asking you to consider the 'journey' the data will/is likely to take within your system, project or process during it's lifetime, you will be able to glean valuable insight into where there may be potential risks that you will need to mitigate against.* . Beneath that, are the following column headers:

Birth Where was the data/form born, i.e. where does the data originate from? who created it and why?

Intended uses/purpose What is the data/form intended for? What is the purpose of the data collection?

Actual uses What is the data/form actually used for? (this may differ from the original intention)

Regularity How often is the data/form used?

Who Who collects the data/fills in the form?

When In what circumstances is the data collected and/or form completed?

Why Why is the data collected and/or form completed?

Format What format is the data/form in (e.g. manual paper based or electronic);

Home (storage) Where does the data/form live after it has been collected and/or when it is stored?

Storage Format In what format is the data/form kept or stored (e.g. raw format, anonymised etc. please explain)

Access Who can access the form once it is stored?

Retention How long is the data/form stored for and why that time period?

Disposal How does the data/form get disposed of?

Transmission does the data/form get moved from it's home (if yes, then please complete a journey for its travels (one for each journey)?

Journey What happens to the form next (this is intended to capture the data journey i.e. whenever the data travels, who it is shared with, how it is shared (faxed, emailed, verbal etc.) and how often this happens. To this end the questions that follow are cyclical and apply every time the form/data (and therefore, the data, travels)). Practitioners are asked to complete the questions for each journey that form/data takes whether that is physical as in the paper being moved, electronic as in the data being shared either by scanning, copying or sending the original. Thus, what happens to the form when when It gets completed or it travels:

1. Where does it go?
2. Who is present?
3. Who sees the data/form?
4. Who can access the data/form?
5. How long does the data/form stay there?
6. Temporary Home - Storage format: If the data/form stays at it's temporary home, in what format is it kept or stored?
7. Temporary Home - Access: Who can access the data/form once it is stored at it's temporary home?
8. Temporary Home - Retention: How long is the data/form stored at its temporary home?
9. Temporary Home - Disposal: How does the data/form get disposed of from it's temporary home?
10. Where does the data/form go next? (this may be multiple places - please complete a journey for each trip)

F.3 PACT Step 2 - Need for a DPIA Questions & Guidance

The instructions provided at the top of this step are: *This section asks you to consider whether a Data Protection Impact Assessment is required for the processing of the data.*
A DPIA is required for any "high risk" processing activities involving personal data. This is a short assessment with drop-down lists that will help you determine whether or not a DPIA is required for your system, project or process.
A YES answer here will mean conducting a DPIA as part of continuing through the rest of the framework (the DPIA assessment is in Step 4).

A NO answer, will and ask you to provide an explanation and the reasons for why a DPIA is not considered necessary. In this case, Step 4 can be skipped. However, it is advised that a DPIA is conducted as a matter of course for any data processing of personal data.
Once complete please move to Step 3, the start of the Risk Assessment.

This questions asked are a the same as those used in ”the need for a DPIA” assessment in the DDW. However, for completeness, they are repeated here:

D - Data Will a new system, project or process be implemented that involves collecting, processing, transmitting, sharing and/or storing of data or are there any changes to an existing system, project or process? (yes/no answer can then be selected from the drop-down menu).

P - Protection Will the system, project or process involve any of the following: (an answer can then be provided from the list of lawful basis for processing shown on the drop-down menu. These can be found in Chapter 2, Section 2.4.4.

Guidance provided If any of these apply DPIA required.

I - Impact Is the system, project or process likely to involve any of the following: (an answer can then be provided from the list of lawful basis for processing shown on the drop-down menu. These can be found in Chapter 2, Section 2.4.4.

Guidance provided If any of these apply DPIA advisable.

Assessment Based on the answers above, DPIA required? (yes/no answer can then be selected from the drop-down menu).

Decision If decision is negative, please record the reasons for not carrying out a DPIA here: (In case of a no answer, a space is then provided for practitioners to note the reason and justification for why a DPIA is not required).

F.4 PACT Step 3 - Wider Context Questions & Guidance

This worksheet is introduced as follows: *This section begins the Risk Assessment. These questions relate to the surrounding context of the activity/scenario/project/system being assessed.*
It is recommended that you complete this section in collaboration with your Privacy Officer an/or Data Protection Officer (DPO) to ensure all contextual considerations are fully considered and covered.
Each of the questions within this section should be considered in relation to all of the actors, the data- subject(s), originator, controller or processor; the organisation or any other related individual whose privacy might or could be affected.
In completing these questions, please refer to the answers provided in the overview, the Data Register and the Data Journey and collected as part of the assessment, to inform the deliberation and help guide the answers you provide.
If a DPIA is required please ensure that in approaching each question, you provide an answer from BOTH the perspective of the Data Subject AND the Organisation in the answer columns provided.

In this worksheet, two additional columns have been added. First, because context is not always an easy concept to imagine, it tends to be something we do as a matter of course and therefore, may not fully appreciate, a column has been provided with some guidance for what the question relates to in terms of context . Second, a column has been added to depict what the norm or value being considered is.

The questionnaire for this Step can be found in Chapter 8, Section 8.4.5). However, for completeness, these have also been repeated below:

Q51 How could data collection, processing, transmission or publication be perceived to infringe on any political values? (please also explain whose values and how they might be infringed upon)

Norm/Value Q51 Political

Guidance Q51 Values are the underlying standards and specific guidelines that we apply to the norms. They are standards or principles of behaviour i.e. our judgement of what is important. Our values are what tells us what is good or bad, what is acceptable and what is not. Values include democracy, freedom of speech and autonomy. In terns of data, our values, societal and individual, will influence how we approach and handle the data which may, for example, be in the best interest for profit but not for the individual whose data is being processed.

Q52 How might data processing be perceived to impose any form of power imbalance? (please explain how such a power balance might arise, who might be affected and how these might be imposed)

Norm/Value Q52 Power imbalance

Q53 How could data processing infringe on any legal obligations? (please clarify what legal compliance might be breached, how and who would be affected)

Norm/Value Q53 Legal

Guidance Q52-53 Values: What values could external or internal stakeholders potentially abuse or be perceived to abuse? Are there any factors, influences or aspects that could be perceived to support/discourage data collection, processing, transmission or publication e.g. what might be the potential effects on implications for justice, power structures, threat to democracy etc.

Q54 How might data processing be perceived to infringe upon any social values? (please explain what those values are, whose values they are and how these might be infringed upon)

Norm/Value Q54 Social

Q55 How might data processing be perceived to infringe any moral/ ethical values? (please explain what those moral values are, whose values and how these might be infringed upon)

Norm/Value Q55 Moral/Ethical

Guidance Q54-55 Are there any ethical factors that may support/discourage data processing e.g. would data collection, processing, transmission or publication threaten fairness, equality or social hierarchy

Norm/Value Q56 Belief system(s)

Q57 How might data processing pose a threat to the autonomy or freedom of the data subject(s)? (please explain how this might pose a threat, who could be affected and how)

Norm/Value Q57 Autonomy & Freedom

Q58 In what ways could data processing result in informational harm on the data subject(s)? (please clarify how such harm could occur, who would be affected and how)

Norm/Value Q58 Harm

Q59 How might data processing give rise to any form of discrimination (please explain what form of discrimination, how this might occur and who would might be adversely affected)

Norm/Value Q59 Discrimination

Q60 In what ways might data processing grant or afford any privileges or prerogatives to the referrer, the data- processor(s), controller(s) or originator(s) in a manner that benefits or is perceived to benefit those parties as a result of data processing?

Norm/Value Q60 Privileges or Prerogatives

Q61 In what ways might data processing result in breach of trust (if yes, please clarify how such harm could occur, who would be affected and how)How could data processing impose security risks on any stakeholders (please clarify what the security risk is, how and who would be affected)

Norm/Value Q61 Security

Q62 In what ways might data processing result in breach of trust (if yes, please clarify how such harm could occur, who would be affected and how)

Norm/Value Q62 Trust

Q63 In what ways could the data processing result in a breach of confidentiality? (please clarify how such a breach might occur, who would be affected and how)

Norm/Value Q63 Confidentiality

Q64 Is there a reasonable expectation on the part of the data subject(s) and/or data controller(s) of data being kept confidential and not collected, processed, transmitted or published?

Norm/Value Q64 Confidentiality

Guidance Q56-64 Norms are customs or expected rules, expectations or standards for how we are supposed to behave in society. Norms can relate to behaviours, ideas and beliefs etc. that have become embedded within us as individuals and society, i.e. influences and standards of expected behaviours or customs. Norms affect how we react, behave and perceive the world around us. Examples of norms include political or religious beliefs, an expectation of non-discrimination, equality (e.g. equal opportunity etc.) Norms: What norms could external or internal stakeholders potentially infringe upon or be perceived to infringe upon?

Q65 What is/are the desired effects of the processing for the data subject?

Q66 What are the benefits and positive values for the data subject that processing the data will bring to enhance the outcome for the data subject e.g. to facilitating treatment, supply goods, improve mobility, increase transparency etc.

Norm/Value Q65-66 Data Subject

Q67 What is the desired effect of the processing for the organisation?

Q68 What are the benefits and positive values that data processing will bring/enhance for the organisation. These may include commercial gain, improved transparency, meeting legal obligation etc.

Norm/Value Q67-68 Organisation

Guidance Q65-68 Positive Impact Assessment: note any positive values that could be derived from the data collection, processing, transmission or publication?

Q69 If consent has not been obtained from the data subject, or they have withdrawn their consent, are there any overriding legal, moral or ethical reasons why data processing should be allowed despite this? (e.g. prevention of terrorism, safeguarding of the data subject etc.)

Q70 Are there any overriding legal, moral or ethical reasons why data processing should be allowed even if there is a risk of re-identification?

Norm/Value Q69-70 Organisation

Guidance Q69-70 Overriding reason for processing despite risks or lack of consent.

F.5 PACT Step 4 - Risks to the Individual

The introduction to this section reads *This section is part of the risk assessment. Please complete and score each risk identified in the risk table below following the guidance in the risk scoring table. Here you are asked to consider what the consequences of the collection, processing, storing, transmitting or sharing of the data associated with this activity/scenario/project/system are likely to be for THE INDIVIDUAL (the data subject).*
i.e. how could this activity/scenario/project/system negatively affect the data subject, what is the likelihood the risk would occur and what would be the severity and impact on the data subject if the risk occurred?
The first section provides some details about the types of risks and where the risk might originate from to help the things you might wish to consider as part of the deliberations and discussions. This is not an exhaustive list, merely a suggestion for the types of risks that might occur.
It is recommended that you complete this section in collaboration with all stakeholders involved to ensure a comprehensive list of privacy risks are identified. Once complete please also consult your Privacy Officer an/or Data Protection Officer (DPO) for a review of the privacy risks identified. Once complete, please move to Step 5 to determine what the risks to the organisation before moving on to PLAN and consider how the risks identified below can be minimised by the organisation.

Some background Information to consider in identifying the privacy risks to the data subject is then provided, these are listed in Table F.2.

Then in Table F.3, the risk scoring is explained. While most scoring matrixes have been devised to look at scoring risks to the organisation (e.g. (NIST 2012), (Lyon and Popov 2016a), (Heiser 2008)), there is one that looks at applying scoring levels to the individual. This scoring matrix has been created by the French Data Protection Authority. (CNIL), as part of their guidance (CNIL

Listed here are some types of breaches that could occur that could have an impact on the data subject (individual).	Types of Privacy Breach that could affect the individual (”Feared Event”)	Unauthorised access to data; Unwanted/unauthorised data modification; Inappropriate use of data; Data damaged; Lost or stolen data; Data interception; Observation of data not authorised to see; Unwanted/unauthorised deletion of data; Unwanted/unauthorised obfuscation of data (e.g. for ransom or due to malfunction)
These will typically originate from an instigator who purposely, unknowingly or unintentionally instigates or carries out the act that causes the breach.	Potential Breach Instigator (”Actor/Risk Source”)	Internal stakeholder (e.g. staff); Internal ’malicious’ stakeholder (e.g. disgruntled staff member); External ’friendly’ stakeholder (e.g. data processor); External ”malicious” non-stakeholder (e.g. hacker, disgruntled customer etc.); Machine / IT System ”non-malicious” (e.g. system failure, unsecure transmission etc.)
In order to facilitate the breach the instigator may use a tool such as an application to aid him/her in carrying out the breach. These will often be the assets that need protecting to minimise the risk.	Tool/method used to facilitate breach (”Supporting Assets”).	Hardware; Software; Cloud based software; Sensory devices; Surveillance cameras; IoT applications; mobile devices; computer channels; USB devices; Humans; Paper documents; paper transmission channels.

Table F.2: Risks to Individuals - Examples to consider

2012). Because the focus is on the individual rather than the organisation, this has been has used as the basis for scoring privacy risks in this section. Further, CNIL provide comprehensive explanation and an easy to follow overview for how each score level applies to the individual, this has therefore been adapted for use here.

The next row then reads *In light of your answers in steps 1 and 3 and the above information, what are the risks to the individual (data subject), please list all that apply in the risk register below.*. This is followed by the following column headings:

Risk ID The risks have been numbered DS1...* However, practitioners will have the option to change the numbering format to suit their requirements;

Risks Identified please note each risk identified on a separate row. If necessary, add additional rows to the spreadsheet;

Risk Likelihood likelihood of harm;

Risk Severity Score- Physical for the Data Subject i.e. likely physical impact on data subject (drop-down selection for the scores; negligible, limited, significant or maximum);

Risk Severity Score - Material for the Data Subject i.e. likely material impact on data subject (drop-down selection for the scores; negligible, limited, significant or maximum);

Risk Severity Score - Ethical for the Data Subject i.e. likely moral/ethical impact on data subject (drop-down selection for the scores; negligible, limited, significant or maximum);

Mitigation Please note any mitigation strategy (in place or planned) that will reduce or eliminate the risk;

Overall Risk Score Taking all the scores into account for each risk - please rate the risk overall (drop-down selection for the scores; negligible, limited, significant or maximum);

Action taken a space is provided to note any action taken to reduce or eliminate the risk;

Date to record the date the action was taken;

Residual Risk (Data Subject (DS)) Please note here any residual risk left post-mitigation and what action will be taken to manage this (drop-down selection for the scores; negligible, limited, significant or maximum);

Risk Scoring Table - Risk Severity (Data Subject) (score levels align with the CNIL Risk Management framework) (CNIL 2012)

Severity concerns the magnitude of a risk and can be estimated based on the extent of potential impacts on data subjects, allowing for safeguards in place (existing or planned controls (these should be mentioned in the mitigation section to help justify the selected score).

Below is a description and examples of what each severity score means for the data subject (adapted from the CNIL Risk Management framework). This table has been included to assist you in thinking about what the consequences and impact of each risk might be for the data subject.

Score Level	Description of impact	Physical Impact. Examples of potential physical impacts e.g. disfigurement, loss of amenity or financial loss related to physical integrity	Material Impact. Examples of potential material impacts e.g. losses incurred with respect to the data subject's assets such as lost revenue	Ethical/Moral Impacts. Examples of potential moral impacts e.g. emotional or physical disfigurement, suffering or loss of amenity.
Negligible	Data subject inconvenienced, but these could be overcome relatively easily.	Dependent/vulnerable person not receiving adequate care; Occasional headaches.	Time loss e.g. having to repeat action; receiving unwanted email (e.g. spam); targeted with adverts (e.g. social network tracking, spam mail).	Invasion of privacy without real harm (e.g. intrusion from commercial actor); Fear of not having control over own personal data; Lack of respect for freedom e.g. to surf the net (e.g. denied access due to controls such as age restrictions); Annoyance: caused by data received or requested, having to spend additional time managing their data; feeling their privacy has been invaded although no real harm has occurred.
Limited	Data subject inconvenienced somewhat but these could be overcome with a few difficulties.	Inadequate care resulting in minor but real harm (e.g. disability); Minor physical complaints (e.g. minor illness caused by disregard of contraindications); Defamation resulting in psychological or physical retaliation.	Financial: unexpected costs (e.g. erroneous fines), additional costs (e.g. interest, fees or charges incurred), payment defaults; Access Denial to services (administrative and/or commercial); Missed/loss of opportunities: e.g. for promotion or of comfort (e.g. cancellation of purchase, leisure or holiday or termination of online account); online account stoppage or blockage; receiving unwanted, targeted mailings causing reputational damage; cost increase (e.g. rise in insurance premiums); data not updated (e.g. previous position recorded); Incorrect data being processed e.g. malfunction in bank account etc.; Targeted advertising related to private or confidential matters which the data subject does not want published (e.g. adverts concerning rehabilitation treatment or pregnancy); inappropriate or inaccurate profiling.	Hindrance/stopped from participation to information systems (e.g. due to whistleblowing, social networks); Minor psychological ailments that could cause harm (defamation, reputation); Relationship problems, personal or professional (e.g. reputational or image damage, loss of status); Feeling of invasion of privacy without irreversible damage; Intimidation (e.g. through social media).
Significant	Significant consequences for the data subject that they should be able to overcome but with serious and real difficulties.	Serious physical complaints resulting in long-term harm (e.g. health worsening due to inadequate care or disregard of counter signs); Adjustment of physical integrity e.g. after an assault, an accident either in the workplace or at home etc.	Financial: loss of money due to fraud (e.g. from phishing attack); Misappropriation of funds without being compensated; ongoing financial difficulties (e.g. requirement to take a loan), Blocked while overseas; Opportunity: loss of one-off, targeted opportunities (e.g. refusal of studies, banned from examination, internship, employment or loan); Prevention: from holding a bank account, loss of home, housing or employment; Property damaged; divorce or separation; Loss of customer data.	Serious psychological condition (e.g. phobia or depression); Feeling of invasion of privacy with irreversible damage; Feeling of vulnerability (e.g. resulting from a court summons); Feeling fundamental rights have been violated (e.g. freedom of expression, discrimination); Becoming a victim (e.g. of blackmail or bullying and harassment (e.g. cyber).
Maximum	Severe and/or irreversible consequences suffered by the data subject that they are unlikely to overcome.	Death (e.g. fatal accident, suicide or murder); Permanent or long-term physical impairment or condition (e.g. due to disregard of counter signs).	Financial: Substantial debts; inability to relocate and/or work; loss of evidence needed for litigation; loss of supply of vital infrastructure supply (electricity, water).	Permanent or persisting psychological conditions; criminal conviction and/or penalty; Abduction; Loss of kinship ties; Inability to litigate; loss of legal autonomy (e.g. guardianship) and/or change of administrative status.

Risk Likelihood - Risk Scoring Table i.e. the likelihood that the risk will occur. Below is a description and examples of what each likelihood score means for the data subject (adapted from the CNIL Risk Management framework).

Score Level	Risk Likelihood (Likelihood of harm)
Negligible	It appears highly unlikely that the selected risk will materialise (e.g. files concerned are stored on a unconnected network, located in a secure room protected by encryption and a lock, badge reader and access code).
Limited	It would be difficult for the risk source to materialise the threat by exploiting the properties of supporting assets (e.g. files concerned are encrypted, not transmitted and stored in a secure network protected by strong safeguards).
Significant	It is possible that the risk will materialise in light of the systems and/or stakeholders involved who could seek to exploit a vulnerability using one or more of the tools available (e.g. files are stored in offices that cannot be accessed without first getting into the network).
Maximum	It is highly likely that the risk could materialise given the system and/or stakeholders connected or likely to be connected, they/it could relatively easily exploiting the/a vulnerability using available tools (e.g. theft of files stored in the open or unsecured repository).

Table F.3: Data Subject Risk Scoring Table

Residual Risk (DS) Likelihood score post-mitigation (drop-down selection for the scores; negligible, limited, significant or maximum);

Residual Risk (DS) Overall Risk Score Taking all the scores into account for each risk - please rate the post-mitigation risk overall. (drop-down selection for the scores; negligible, limited, significant or maximum).

F.6 PACT Step 5 - Risks to the Organisation Questions & Guidance

The introductions provided at the beginning of this assessment state: *This section forms the final part of the risk assessment. Please complete and score each risk identified in the risk table below following the guidance in the risk scoring table.*
Here you are asked to consider what the consequences of the processing of the data associated with this activity/scenario/project/system are likely to be for THE ORGANISATION
i.e. how could this activity/scenario/project/system negatively affect the organisation, what is the likelihood the risk would occur and what would be the severity and impact on the organisation if the risk occurred?
The first section provides some details about the types of risks and where the risk might originate from to help the things you might wish to consider as part of the deliberations and discussions. This is not an exhaustive list, merely a suggestion for the types of risks that might occur.
It is recommended that you complete this section in collaboration with all stakeholders involved to ensure a comprehensive list of privacy risks are identified. Once complete you may wish to also consult your Privacy Officer an/or Data Protection Officer (DPO) for a review of the privacy risks identified.
Once complete, please move to Step 6 to PLAN and consider how the risks identified below can be minimised by the organisation.

The next section provides some *Background Information to consider in identifying the privacy risks to the Organisation* and then the following guidance is provided:

The aim of this assessment is to prevent disclosure of data that the organisation does not want to share or disclose for one of the following reasons:

Legal Obligation Data that is protected by law, i.e. any personal data which is protected under data protection regulations such as GDPR;

Commercial sensitivity Referring to any commercially sensitive data the organisation wishes to protect;

Policy Any data that organisational policy requires be kept confidential;

Confidentiality Any data subject to confidentiality clauses e.g. under contract or non-disclosure agreements;

Other any other industry specific reason that can justifiably be applicable.

Listed here are some types of breaches that could occur that could have an impact on the Organisation

Types of Privacy Breach that could affect the individual (”Feared Event”) Unauthorised access to data; Unwanted/unauthorised data modification; Inappropriate use of data; Data damaged; Lost or stolen data; Data interception; Observation of data not authorised to see; Unwanted/unauthorised deletion of data; Unwanted/unauthorised obfuscation of data (e.g. for ransom or due to malfunction).

These will typically originate from an instigator who purposely, unknowingly or unintentionally instigates or carries out the act that causes the breach.

Potential Breach Instigator (”Actor/Risk Source”) Internal stakeholder (e.g. staff); Internal ’malicious’ stakeholder (e.g. disgruntled staff member); External ’friendly’ stakeholder (e.g. data processor); External ’malicious’ non-stakeholder (e.g. hacker, disgruntled customer etc.); Machine / IT System ’non-malicious’ (e.g. system failure, unsecure transmission etc.).

In order to facilitate the breach the instigator may use a tool such as an application to aid him/her in carrying out the breach. These will often be the assets that need protecting to minimise the risk.

Tool/method used to facilitate breach (”Supporting Assets”) Hardware; Software; Cloud based software; Sensory devices; Surveillance cameras; IoT applications; mobile devices; computer channels; USB devices; Humans; Paper documents; paper transmission channels.

This is followed by a risk scoring table (see Table F.4) similar to the one used in Step 4. The scoring levels have been kept the same for completeness and to avoid having too many different scoring levels to consider. For Step 5 however, the following changes have been made:

1. The examples provided have been adapted to consider the organisation;
2. The severity level for ethical/moral impact has been replaced with ”reporting/notification implications”.

Below Table F.4 is some suggested questions for practitioners to consider. The heading to this section reads: *Below are some suggested questions you may wish to ask. This is not an exhaustive list, merely a suggestion for the types of risk you may consider as part of the deliberation and discussions.*. This is followed by the following list of questions:

Asset disclosure risk Questions to think about. Please consider the information from the previous steps in creating the list of risks. In particular make sure that any risks identified in Step 4 (risks to the individual) are addressed as part of this exercise.

- In what ways might data be disclosed to unauthorised parties?
- In what ways could data be accessed by unauthorised parties?
- In what ways might data be deleted unintentionally or maliciously, how would this affect operations?
- In what circumstances might data be modified unintentionally or maliciously, how would this affect operations?

Risk Scoring Table - Risk Severity (score levels align with the CNIL Risk Management framework) (CNIL 2012)

Severity concerns the magnitude of a risk and can be estimated based on the extent of potential impacts on the organisation, allowing for safeguards in place (existing or planned controls (these should be mentioned in the mitigation section to help justify the selected score). Below is a description and examples of what each severity score means for the organisation (adapted from the CNIL framework). This table has been included to assist you in thinking about what the consequences and impact of each risk might be for the organisation.

Score Level	**Description of impact**	**Physical Impact** Examples of potential physical impacts e.g. hardware destroyed, employee harmed	**Material Impact** Examples of potential material impacts e.g. losses incurred with respect to assets such as lost revenue	**Ethical/Moral Impact** Examples of potential moral impacts e.g. the risks identified are against policy, the company ethos etc.	**Reporting/ notification implications.**
Negligible	Organisation inconvenienced, but this can be overcome relatively easily.	Data or system failing leading to a manual restart in a different location being required, no loss of data.	Time loss e.g. having to travel to location, slight downtime of system; Inability to access system and/or data for short time period, business relatively unaffected; annoyance caused by having to spend additional time managing the system or data; no real harm has occurred.	Invasion of data subject (DS) privacy without real harm (e.g. pop-up adverts); Denying DS access to services or website for false reason (e.g. assumed under age when that is not the case); Annoyance: Sending promotional emails to customers too often (e.g. multiple times daily).	No need to report incident beyond operational level.
Limited	Organisation inconvenienced somewhat but this can be overcome with a few difficulties.	Data or system failing leading to a having to purchase minor new equipment; loss of access to data for longer period but with little effect on the business, no loss of data.	Financial: unexpected minor costs (e.g. new equipment), additional costs (e.g. interest, fees or charges incurred); online account stoppage or blockage; data not updated (e.g. sales during downtime not recorded); Incorrect or historical data being used for processing while system is down etc.	Refusing DS access to participation in forum discussion (e.g. due to whistleblowing, social networks); publishing or sharing incorrect data (defamation, reputation); Sending spam emails to customers not signed up to receive such mailings; Delayed response to a DS request for access under GDPR (e.g. delaying or refusing to update incorrect record).	Incident has to be reported to mid-level management.
Significant	Significant consequences for the organisation that are recoverable from but with serious and real difficulties.	Data or system failing leading to a having to purchase major new equipment; significant delays/downtime of system resulting in staff or stakeholder loss of access to premises, equipment and/or data for longer period affecting the business, data loss.	Brand, reputational or image damage; loss of custom(ers); Financial: loss of money due to fraud (e.g. from phishing attack); Misappropriation of funds; Opportunity: loss of one-off, targeted opportunities (e.g. loss of sale); Property damaged; Loss of customer data; fine from data protection authorities.	Staff processing or sharing data contrary to organisational policy e.g. without consent being obtained; Using automated decision making algorithms that result in unintentional discrimination or bias (e.g. targets one area or ethnical group as more problematic than another); Targeting minority group for commercial gain without due regard for privacy or legitimacy of singling this group out; Incorrectly charging customers, taking significant sums from their account; Refusing a DS their rights under GDPR (e.g. refusing the right of erasure); Not responding or addressing reported discrimination, blackmail or bullying on organisational website.	Senior management and DPO has to be kept informed. Data protection authority need to be notified of breach.
Maximum	Severe and/or irreversible consequences suffered by the organisation that will be irrecoverable or take substantial resource and effort to be overcome.	Death (e.g. fatal accident, suicide or murder of staff member); Permanent or long-term damage to equipment or other assets.	Substantial: costs, Brand, reputational or image damage; loss of custom(ers); misappropriation of funds; loss business opportunities (e.g. loss of major customer or contract); Property damaged; Loss of customer data; fine from data protection authorities, loss of supply of vital infrastructure supply (electricity, water).	Placing personal data in open repository online; changing DS administrative status incorrectly and not rectifying this; litigating against DS using false information to get a conviction; data breach e.g. employee losing, selling/stealing personal data to unauthorised third-parties; Using automated decision making algorithms that result in direct discrimination or bias (e.g. targets one area or ethnical group as more problematic than another).	Senior management and DPO has to be kept informed. Data protection authority need to be notified of breach.

Risk Likelihood - Risk Scoring Table i.e. the likelihood that the risk will occur. Below is a description and examples of what each likelihood score means for the organisation (adapted from the CNIL PIA framework).

Score Level	**Risk Likelihood (Likelihood of harm)**
Negligible	It appears highly unlikely that the selected risk will materialise (e.g. files concerned are stored on a unconnected network, located in a secure room protected by encryption and a lock, badge reader and access code).
Limited	It would be difficult for the risk source to materialise the threat by exploiting the properties of supporting assets (e.g. files concerned are encrypted, not transmitted and stored in a secure network protected by strong safeguards).
Significant	It is possible that the risk will materialise in light of the systems and/or stakeholders involved who could seek to exploit a vulnerability using one or more of the tools available (e.g. files are stored in offices that cannot be accessed without first getting into the network).
Maximum	It is highly likely that the risk could materialise given the system and/or stakeholders connected or likely to be connected, they/it could relatively easily exploiting the/a vulnerability using available tools (e.g. theft of files stored in the open or unsecured repository).

Table F.4: Organisational Risk Scoring Table

- In what ways might data be stolen or removed unintentionally or maliciously, how would this affect operations?
- In what ways might data be obfuscated or encrypted with no key making it unaccessible unintentionally or maliciously, how would this affect operations?
- What are the risks of the data being transmitted, shared or published to the wrong recipient?
- Is data subject(s) aware of data being disclosed (processed/shared/published), if not how might this affect operations?
- Lack of consent from the data subject to data collection/processing/disclosure how might this affect the organisation?
- Lack of consent from the data subject to secondary data collection/processing/disclosure how might this affect the organisation?
- What are the risks of the attributes, once transmitted or published, being linked to external data in such a way that they can pose a risk of contributing to re-identification of a data subject? (please be specific as to what risk and how this might pose a new risk)
- In light of the answers provided to the wider context questions in Step 3, were any risks identified that could affect operations?
- Identify any system access control processes that may be relevant and discuss how these might be vulnerable?

This if followed by the actual risk register which mirrors the risk register in Step 4 with the focus on the organisation rather than the individual. Therefore, the headings follow the risk scoring table in Table F.4. Further the risk ID's have been numbered O1...* to reflect the organisational perspective.

F.6.1 PACT Step 6 - PLAN

Step 6 of PACT, has been designed around the 7-DL, used to frame the Privacy Lifecycle PlAN (PLAN). This consist of a series of questions devised to help practitioners consider how they will manage the data throughout its lifecycle with the organisation. This worksheet in the spreadsheet consist of the following column headers:

Q No. the question numbers, (currently pre-set with 7DS1 - 7DS64);

Data Lifecycle Stage (7-DS) No. (see Figure 8.7);

Description of 7-DS Stage these can be found in Section 8.5.1;

Relevant GDPR Principle(s) (see Table F.5);

Guidance See Table F.5;

QUESTIONS *Referring to the answers collected as part of the assessment, please answer the following questions. These are designed to help with forward planning for how the data will be managed during its lifecycle* (see Table F.5).

Answers Providing a space for practitioners to answer each question;

Review - answers (if different from original) providing a space for future review of the PLAN;

Review completed by ; and

Date of Review.

This step is introduced as follows: *This section has been included to assist you in creating a Privacy Lifecycle PlAN (PLAN). It has been included to help you plan how the data will be managed during its life time with the organisation i.e. the data lifecycle. The PLAN is designed around a 7-step Data Life Cycle (7-DS) to encourage you to consider privacy throughout its life time with the organisation. It has been used to delineate the different stages of data: collection, transformation, retention, access/release, post-access, disposal and governance and consultation. In addition, some ideas for consultation have also been provided on the information worksheet at the beginning of the framework.* This is followed by a series of questions and guidance as listed in Table F.5.

Q No.	7-DS Stage	Relevant GDPR Principle(s)	Guidance	QUESTIONS
7DS1	1	Legality of collecting/ processing data/Transparency		What is the scope of the data collection/processing - will we require consent?
7DS2	1	Legality of collecting/ processing data	GDPR places a requirement on organisations to specify a lawful basis for processing of personal data	What is the primary lawful basis for collecting/processing or handling the data? (please refer to Q3 on the overview worksheet to determine which lawful basis is relevant for your data)
7DS3	1	Legality of collecting/ processing data/Transparency	Where special category data (e.g. data relating to health) is processed, a secondary lawful basis for processing may also be selected e.g. the primary lawful basis may be contractual with consent as a secondary lawful basis for certain elements.	If applicable, what is the secondary lawful basis for collecting/processing or handling the data? (please refer to Q4 on the overview worksheet to determine which secondary lawful basis is relevant for your data)
7DS4	1	Proportionality/ Fair data processing/ collection	Proportionality requires that any limitation on the rights of the individual have to be justified. For example, making sure that the measure(s) taken in processing the data do not disproportionally limit the rights of the individual whose data is being processed. A pre-condition is that the measure(s) taken in processing or safeguarding are sufficient to achieve the objective while only relevant personal data for the purposes of the processing is collected and processed.	What is the purpose of collecting/processing the data?
7DS5	1	Proportionality/ Fair data processing/collection	as question 7DS4	How will the data be collected?
7DS6	1	Proportionality/ Fair data processing/collection	as question 7DS4	Please note what the reason for collecting, processing and/or sharing the data is?
7DS7	1	Data Minimisation/ proportionality	as question 7DS4	How much data will be collected? Can this be reduced to minimise risk?
7DS8	1	Data Minimisation/ proportionality	as question 7DS4	Has data limitation/minimisation and proportionality been considered, is all the data being collected necessary?
7DS9	1	Purpose limitation/ relevance	as question 7DS4	Has/can a justification be made for each item of data being collected?
7DS10	1	Transparency	as question 7DS4	If personal data is collected/processed has data subject been informed how the data will be used and their rights?
7DS11	2	Proportionality/Fair data processing/collection	Ensuring the data is updated, corrected as required and accurate	How will the data be processed and used (processed)?
7DS12	2	Confidentiality & Integrity/fair data processing	as question 7DS11	How is/will the data be accessed for use/processing?
7DS13	2	Accuracy	as question 7DS11	How will the quality and accuracy of the data be maintained?
7DS14	2	Accuracy	as question 7DS11	Who is responsible for data quality?
7DS15	All		Privacy Goal: Confidentiality - Ensuring data is only accessible to authorised stakeholders	What procedures and measures are in place to safeguard that confidentiality is achieved and maintained?
7DS16	All	Confidentiality	Privacy Goal: Integrity - Ensuring non-repudiation and reliability for each piece of data, i.e. processing correct, authentic and unmodified data	How will you protect the accuracy or completeness of this data against unintended modification?

Continued on next page

7DS17	All	Confidentiality & Integrity	Privacy Goal: Availability - Ensuring data is usable on demand and accessible to authorised stakeholders	What procedures and measures are in place to ensure the data is accessible, comprehensible, and processable when authorised stakeholders want to use or access it?
7DS18	All	Confidentiality & Integrity	Privacy Goal: Unlinkability - Ensuring data cannot be sufficiently distinguished to linked across platforms or domains with similar context or purposes	How will you ensure data cannot be linked across entities, platforms or domains?
7DS19	All	Confidentiality & Integrity	Privacy Goal: Unobservability /Undetectability - Ensuring data is anonymised so that a anonymity and undectability of the individual is preserved	What procedures and measures are in place to facilitate that no individual data subject can be tracked, observed or detected from the data?
7DS20	All	Confidentiality & Integrity	Privacy Goal: Anonymity - Obfuscating links between data and identity i.e. the ability to distinguish any one individual from the data	What measures and procedures are in place to anonymise he data and how will anonymity be maintained and safeguarded?
7DS21	All	Confidentiality & Integrity	Privacy Goal: Pseudonymity - Replacing identifying data with pseudonyms ensuring any links to original data cannot be made by unauthorised parties	What measures and procedures are in place to pseudonymise the data and how will this be maintained and safeguarded?
7DS22	All	Confidentiality & Integrity	Privacy Goal: Intervenability - Enabling data subject access and/or supervisory authority access to affect action on the records (e.g. request modification and/or deletion). In that way it can be seen as a safeguarding measure that must be included within any process or system involving personal data	What is the procedure for responding the data subject access requests, deletion and/or withdrawal of consent?
7DS23	All	Confidentiality & Integrity/ Transparency/ Fair Use	Privacy Goal: Intervenability - (guidance as 7DS22)	What is the procedure for responding the data protection authority request for information and/or access to the data?
7DS24	All	Transparency	Privacy Goal: Transparency - Openness - Providing assurance, accountability and traceability for internal and external stakeholders	What measures are in place to communicate to the data subjects what their rights are?
7DS25	All	Transparency	Privacy Goal: Transparency -GDPR requires that ”any information addressed to the public or to the data subject be concise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualisation be used”	What steps have been taken to ensure such communication is in ”plain and clear language” as required by GDPR?
7DS26	All	Transparency	Privacy Goal: Transparency - (guidance as 7DS25)	Has any communication to the data subject been approved by your DPO (or Data Protection Officer)?
7DS27	All	Transparency	Privacy Goal: Transparency- (guidance as 7DS25)	What measures are in place to communicate to other external parties about procedures and policies in place relating to the data and any processing?
7DS28	All	Confidentiality & Integrity	Access Control/Authorisation	What procedures are in place for setting, maintaining and monitoring data access and processing controls?
7DS29	All	Confidentiality & Integrity	Access Control/Authorisation	Who is responsible for maintaining data access and processing controls?
7DS30	All	Confidentiality & Integrity		Who is responsible for security of manual data handling, processing or storing?
7DS31	All	Confidentiality & Integrity		Who is responsible for security of electronic data handling, processing or storing?
7DS32	3	Storage limitation		How/where will the data be stored?
7DS33	3	Storage limitation		Who is responsible for data storage?
7DS34	3	Storage limitation		Please explain what, if any, procedures will be done to the data in preparation for storage? e.g. If the data is stored, will this be in ’raw’ format or will the data be anonymised or pseudonymised?
7DS35	3	Storage limitation		How long will the data be retained?
7DS36	3	Storage limitation		What is the justification for the retention period selected?
7DS37	3	Storage limitation		Who is responsible for data storage and retention periods?
7DS38	4	Confidentiality & Integrity/Fair Use		Please explain what, if any, procedures will be done to the data in preparation for transmission? e.g. will the data be encrypted prior to/during/after transmission?
7DS39	4	Transparency		Data subject right to access and erasure requests, how is/will this be managed/accommodated?
7DS40	4	Confidentiality & Integrity/ Transparency/ Fair Use		Who is responsible for responding to data subject inquiries/access requests?
7DS41	4	Confidentiality & Integrity/ Transparency		How will we meet other stakeholders expectations?

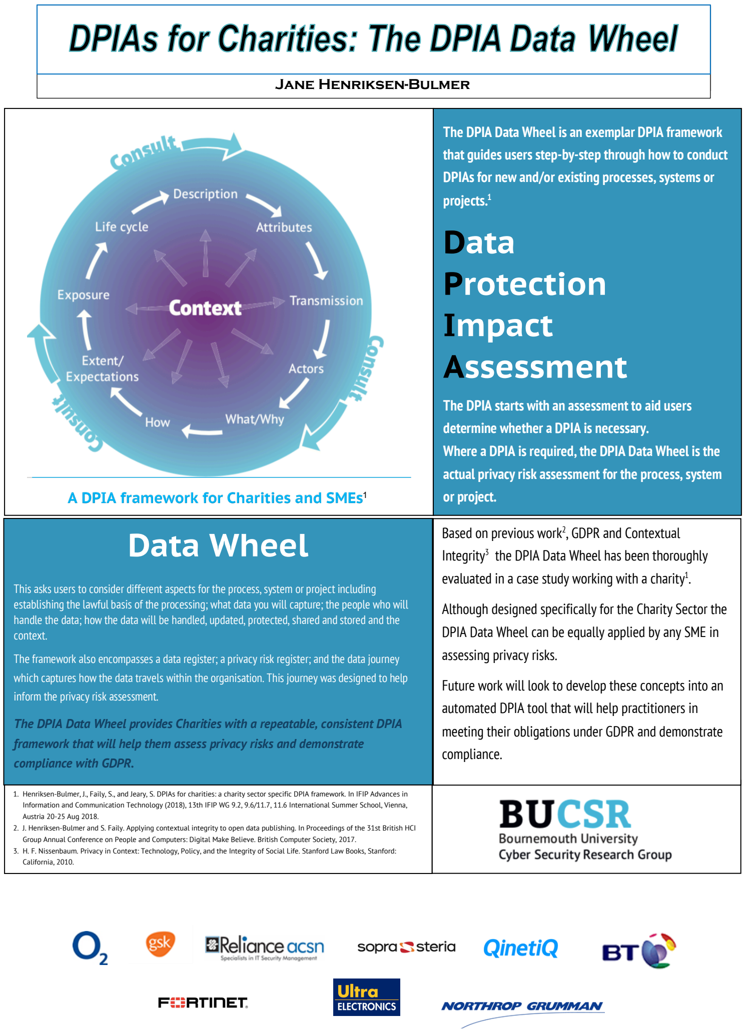
Continued on next page

7DS42	5	Confidentiality & Integrity/ Transparency/ Fair Use	Third-parties refers to any actors/stakeholders who collect, use or otherwise handle the data. These may include the data subject(s) (i.e. the individuals whose data is being collected and/or processed) and any other stakeholders such as an external: data controller (DC), data owner (DO), Data Protection Officer (DPO), Data Processor (DP), or Joint Data Controller (JC). Further details and descriptions are available in the instructions worksheet (under actors)	If any non-contractual third-parties have access to/use of the data how will this be managed?
7DS43	5	Confidentiality & Integrity/ Proportionality/ Transparency/ Fair Use	as 7DS42	How will non-contractual third-party access to/use of the data be monitored?
7DS44	5	Confidentiality & Integrity/ Proportionality/ Transparency/ Fair Use	as 7DS42	If any non-contractual third-parties have access to/use of the data who is responsible for ensuring they do not receive any personal and/or confidential data?
7DS45	5	Confidentiality & Integrity/ Proportionality/ Transparency/ Fair Use	as 7DS42	If any contracted third-parties process/use the data on your behalf (DP and/or JC) who is responsible for ensuring they have appropriate safeguarding measures in place?
7DS46	5	Confidentiality & Integrity/ Proportionality/ Transparency/ Fair Use	as 7DS42	If any contractual third-parties have access to/use of the data who is responsible for ensuring they have appropriate safeguarding measures in place?
7DS47	5	Confidentiality & Integrity/ Proportionality/ Transparency/ Fair Use	as 7DS42	How will contractual third-party DP's and/or JC's processing/use of the data be monitored?
7DS48	5	Confidentiality & Integrity/ Proportionality/ Transparency/ Fair Use	as 7DS42	Who is responsible for auditing third party use of data compliance?
7DS49	5	Confidentiality & Integrity/ Proportionality/ Transparency/ Fair Use	as 7DS42	If a third party suffers a breach involving your data, how will this be managed?
7DS50	5	Confidentiality & Integrity/ Proportionality/ Transparency/ Fair Use	as 7DS42	Who is responsible for collaborating with the third party in implementing the DB plan?
7DS51	6	Storage limitation		How will data be 'retired' and/or disposed of at the end of its lifecycle? Please explain what processes are in place/will be put in place to dispose of/destroy the data?
7DS52	6	Storage limitation		Who is responsible monitoring and auditing data disposal?
7DS53	6	Storage limitation		Who is responsible for data disposal?
7DS54	7	Compliance		How is/will compliance be measured and controlled?
7DS55	7	Compliance		How is/will compliance be evidenced?
7DS56	7	Compliance		Who is responsible for auditing compliance?
7DS57	7	Confidentiality & Integrity/ Proportionality/ Transparency/ Compliance		Is a business continuation (BC) plan in place?
7DS58	7	Confidentiality & Integrity/ Proportionality/ Transparency/ Compliance		Who is responsible for invoking and managing the BC plan?
7DS59	7	Confidentiality & Integrity/ Proportionality/ Transparency/ Compliance		Is a procedure for responding to and dealing with a data breach (DB) in place?
7DS60	7	Confidentiality & Integrity/ Proportionality/ Transparency/ Compliance		Who is responsible for invoking and managing the DB plan?
7DS61	7	Confidentiality & Integrity/ Proportionality/ Transparency/ Compliance		Data subject right to access and erasure requests, is a policy and process for dealing with such requests in place?
7DS62	7	Consultation		Has consultation with stakeholders taken place?
7DS63	7	Consultation		Who is responsible for stakeholder consultation?
7DS64	7	Consultation		How will stakeholder's feedback be incorporated into ongoing practices?

Table F.5: PLAN Questions

Appendix G

DPIA Poster



DPIAs for Charities: The DPIA Data Wheel

JANE HENRIKSEN-BULMER

A DPIA framework for Charities and SMEs[1]

The DPIA Data Wheel is an exemplar DPIA framework that guides users step-by-step through how to conduct DPIAs for new and/or existing processes, systems or projects.[1]

Data
Protection
Impact
Assessment

The DPIA starts with an assessment to aid users determine whether a DPIA is necessary.

Where a DPIA is required, the DPIA Data Wheel is the actual privacy risk assessment for the process, system or project.

Data Wheel

This asks users to consider different aspects for the process, system or project including establishing the lawful basis of the processing; what data you will capture; the people who will handle the data; how the data will be handled, updated, protected, shared and stored and the context.

The framework also encompasses a data register; a privacy risk register; and the data journey which captures how the data travels within the organisation. This journey was designed to help inform the privacy risk assessment.

The DPIA Data Wheel provides Charities with a repeatable, consistent DPIA framework that will help them assess privacy risks and demonstrate compliance with GDPR.

Based on previous work[2], GDPR and Contextual Integrity[3] the DPIA Data Wheel has been thoroughly evaluated in a case study working with a charity[1].

Although designed specifically for the Charity Sector the DPIA Data Wheel can be equally applied by any SME in assessing privacy risks.

Future work will look to develop these concepts into an automated DPIA tool that will help practitioners in meeting their obligations under GDPR and demonstrate compliance.

1. Henriksen-Bulmer, J., Faily, S., and Jeary, S. DPIAs for charities: a charity sector specific DPIA framework. In IFIP Advances in Information and Communication Technology (2018), 13th IFIP WG 9.2, 9.6/11.7, 11.6 International Summer School, Vienna, Austria 20-25 Aug 2018.
2. J. Henriksen-Bulmer and S. Faily. Applying contextual integrity to open data publishing. In Proceedings of the 31st British HCI Group Annual Conference on People and Computers: Digital Make Believe. British Computer Society, 2017.
3. H. F. Nissenbaum. Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford Law Books, Stanford: California, 2010.

BUCSR
Bournemouth University
Cyber Security Research Group

Figure G.1: IAAC Poster