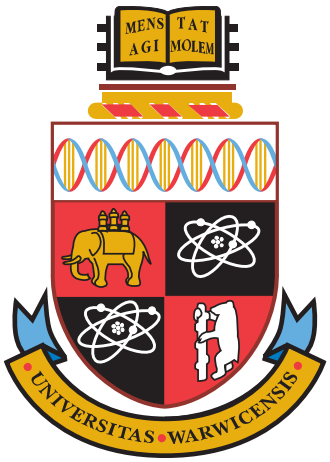


Enhancing user's privacy: Developing a model for managing and testing the lifecycle of consent and revocation

Ioannis Agrafiotis

(BSc, MSc)

A thesis submitted in partial fulfilment of the requirements for

the degree of

Doctor of Philosophy in Engineering

School of Engineering

University of Warwick

2012

Contents

List of Figures

List of Tables

Acknowledgements

In the last years I was fortunate to have the academic and moral support of several people. Without their help and encouragement this work would not have been completed.

First and foremost, I would like to thank my supervisors, Professor Sadie Creese and Professor Michael Goldsmith, for their invaluable guidance, advice and patience. I am grateful for the opportunities they offered me, the encouragement and support in this journey into the academic world. I hope this journey will continue. They made all PhD comics look so alien to me!

I would also like to thank Professor Michael Huth and Professor Yasmin Merali for agreeing to be my examiners and for their valuable comments that strengthen this thesis.

I am grateful for the financial support provided by the EPSRC. I would also like to thank the EnCoRe team for all the valuable input, discussions and contributions to this thesis. In particular, I would like to thank Dr. Edgar Whitley for giving me access to the transcripts of the focus groups. Without his assistance the qualitative part of the research would have been impossible. I would also like to thank HW Communications Ltd. for implementing the testing code. Dr. David Lund, Dr. Bassem Ammar, George Mourakis and Dr. Noel Catterall thank you all very much for your help.

Special gratitude goes to former and current members of the e-Security Group at the University of Warwick, now transferred to the Cybersecurity Group at Oxford. In particular, I would like to thank Dr. Nick Papanikolaou for helping me when I first joined the group and for his ongoing encouragement, motivation and advice throughout these years. I will never forget the long discussions over coffee. Also, I would like to thank Dr. Nick Moffat, who over the last two years offered me valuable guidance and support. His diligence and patience allowed me to improve my work, and as he usually says, "this is music to my ears!". Furthermore, I would like to thank Dr. Jason Nurse, Dr. Jassim Happa, Thomas Gibson-Robinson, Adrian Duncan, Dr. Adedayo Adetoye, Dr. Duncan Hodges, Syed Sadiqur Rahman, Dr. Mark Josephs and Paul Hopkins for their advice, help, encouragement and motivation. Special thanks to Elmedin Selmanovic, Dr. Tom Bashford-Rogers,

Mike Auty and Dr. Konstantinos Bavelis for making easier the long nights spent at the Digital lab and for the joyful moments at the University of Warwick.

I would also like to thank Chrysanthi Papoutsi, Alexandros Georgalis, Elena Pasia, Dimitris Valsamidis for the relaxing moments and their support throughout these years. In particular, I would like to thank for all the fun moments, my friends in Greece: Sokratis Vatalis, Marina Geka, Dimitris Antoniadis, George Michailidis, Apostolis Gouziotis, Lena Dramanidou, Ntinos Fasoulas, George Giannikis, Anastasios Annoulidis, Prokopis Didilis, Athanasios Dislakis, Germanos Gavrilidis, Stella Pasvalidou, Alexandros Ioakeimidis, Despoina Christodoulou, Anastasia Papageorgiou, Fanis Korlos, Dimitris Samaras and George Gerogiannis. I could not have completed my work without your motivation.

Most importantly I would like to express my deepest gratitude to my family. Dad, Mom, Victoria and Nick without your support and direction I would not have achieved anything. Thank you for believing in me.

As you set out for Ithaca
hope that your journey is a long one,
full of adventure, full of discovery.
Laistrygonians and Cyclops,
angry Poseidon-don't be afraid of them:
you'll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare sensasion
touches your spirit and your body.
Laistrygonians and Cyclops,
wild Poseidon-you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.

Hope that your journey is a long one.
May there be many summer mornings when,
with what pleasure, what joy,
you come into harbors you're seeing for the first time;
may you stop at Phoenician trading stations
to buy fine things,
mother of pearl and coral, amber and ebony,
sensual perfume of every kind-
as many sensual perfumes as you can;
and may you visit many Egyptian cities
to learn and learn again from those who know.

Keep Ithaka always in your mind.
Arriving there is what you're destined for.
But don't hurry the journey at all.
Better if it lasts for years,
so that you're old by the time you reach the island,
wealthy with all you've gained on the way,
not expecting Ithaca to make you rich.
Ithaca gave you the marvelous journey.
Without her you would have not set out.
She has nothing left to give you now.

And if you find her poor, Ithaca won't have fooled you.
Wise as you will have become, so full of experience,
you'll have understood by then what these Ithacas mean.
Ithaka, C. P. Kavafis, 1911

Declaration

This thesis is written in accordance with the regulations for the degree of Doctor of Philosophy. The work to be presented has been composed by myself and has not been submitted for a degree at another university. The material in this thesis has been undertaken by myself, except where otherwise stated. Parts of this thesis have been published previously in conferences as follows.

The conceptual model in Chapter 3 has been published in a conference paper in [10]. The formalisations in Chapter 5 have been published in a conference paper in [11]. Work presented in Chapter 6 has been published in a conference paper in [8], while the testing strategy methodology, presented in Chapter 8 has been published for a conference paper in [7].

Ioannis Agrafiotis, August 2012

Abstract

Increasingly, people turn to the Internet for access to services, which often require disclosure of a significant amount of personal data. Networked technologies have enabled an explosive growth in the collection, storage and processing of personal information with notable commercial potential. However, there are asymmetries in relation to how people are able to control their own information when handled by enterprises. This raises significant privacy concerns and increases the risk of privacy breaches, thus creating an imperative need for mechanisms offering information control functionalities.

To address the lack of controls in online environments, this thesis focuses on consent and revocation mechanisms to introduce a novel approach for controlling the collection, usage and dissemination of personal data and managing privacy expectations. Drawing on an extensive multidisciplinary review on privacy and on empirical data from focus groups, this research presents a mathematical logic as the foundation for the management of consent and revocation controls in technological systems.

More specifically, this work proposes a comprehensive conceptual model for consent and revocation and introduces the notion of 'informed revocation'. Based on this model, a Hoare-style logic is developed to capture the effects of expressing individuals' consent and revocation preferences. The logic is designed to support certain desirable properties, defined as healthiness conditions. Proofs that these conditions hold are provided with the use of Maude software. This mathematical logic is then verified in three real-world case study applications with different consent and revocation requirements for the management of employee data in a business environment, medical data in a biobank and identity assurance in government services. The results confirm the richness and the expressiveness of the logic. In addition, a novel testing strategy underpinned by this logic is presented. This strategy is able to generate testing suites for systems offering consent and revocation controls, such as the EnCoRe system, where testing was carried out successfully and resulted in identifying faults in the EnCoRe implementation.

Keywords: Privacy, Consent, Revocation, Formal Methods, Hoare Logic, Testing Strategy

CHAPTER 1

Introduction

There have been inventions in the history of mankind that changed the way people communicate, causing such an impact on societal norms that the world became a different place. For example, the printing press in the 15^{th} century ended the monopoly of Church to knowledge and its control on the ideas that were published in the Western world. The printing press expanded the flow of information to new audiences, giving them the opportunity to be educated and to freely publish their thoughts and ideas. In contemporary times, the Internet has altered our society's function since most of the economic, professional and social aspects of our lives now hugely depend on networked technologies.

Individuals use the Internet to acquire access to products, services and benefits, to articulate their political or religious beliefs, to form their social relationships or operate their businesses. Technological innovations offer opportunities to enterprises for novel and agile business models and ever-growing capabilities to collect, store, process and share huge quantities of personal data in cyberspace. As Francis Maude noted in a speech at the Information Commissioner's Conference on 6 March 2012, "cyberspace has become a vast storehouse for human knowledge" since there is an abundance of data and sophisticated tools to analyse this data [96].

However, the use of online applications comes with compromise. Data subjects, a term that in this thesis will be used to describe individuals whose data is handled by others, when disclosing personal data on the Internet they practically lose any control over how this data is handled. For example, information uploaded by users on social-networking sites is often analysed and sold to enterprises, and users are categorized in profiles according to their demographic data and commercial preferences. Mechanisms to enable users to control these actions are missing and companies are reluctant to implement a system that manages controls and miti-

gates the risk associated with the absence of such controls, for example, allow data subjects to remove or modify personal data held. This lack of controls may hinder data subjects from trusting data controllers, a term defined in this thesis to include all the parties handling and processing personal data. When the data controllers offer their services on cyberspace, the lack of controls raises concerns regarding data subjects' privacy protection.

That these concerns are based on solid ground is illustrated by the increasing number of incidents where data has been lost, mistreated, or shared without authority [1], making the use of privacy-enhancing technologies essential for every Internet user. For example, in 2005 Sony BMG's anti-piracy measure resulted in vulnerabilities and loss of personal data, forcing the company to pay $150 to its customers in Texas [1]. Since 2007, Google has been facing privacy complaints due to the Street View application resulting in a fine of $142,000 by the French government [1], while on 2010, Facebook's most popular applications confessed to sharing users' personal data with advertising companies [1].

The absence of a system able to manage mechanisms to control personal data in online environments does not only raise concerns for data subjects but imposes consequences for data controllers as well. Enterprises have often been reluctant to design and implement such mechanisms due to the financial cost and the constraints that these would impose on enterprise data-handling practices. But with the new EU Privacy and Communications Directive [80] it is estimated that unless businesses implement a mechanism to obtain and manage explicit consent for websites, they could lose up to ten billion pounds [88].

In the recent years, voices acknowledging the need to consider privacy principles when personal information is used by data controllers are increasing and privacy advocates, governments and consumers require companies "to cater for privacy concerns" [124]. Technological developments and the use of the Internet have crafted current business practices to treat personal data as a commodity, without providing any sense of control to the data subjects. However, the importance of information privacy is challenging researchers from different disciplines to propose new information systems that will shift business norms towards privacy-friendly practices. According to Rule, today, information is indefinitely stored, and there is no control regarding the future handling of the data or the context in which it is shared further [223].

There are limited references to control mechanisms in the literature on privacy. Privacy controls have only recently been introduced in large-scale information sys-

tems, and the use of privacy-impact statements is still a maturing discipline. Social-networking sites, such as Facebook and Twitter, include embedded mechanisms to capture users' preferences regarding their consent, which offers some semblance of control, while governmental directives require from companies to clearly state the use of cookies in their websites[129].

However, whilst users may consent explicitly to the sharing, storing and processing of data on such sites, they cannot as easily revoke (permissions to hold or process) data they may already have disclosed. This means that, in most cases, it is not possible for users to change their privacy preferences in a transparent way; without an explicit revocation capability users cannot have clear and unambiguous controls to protect data privacy. Unfortunately, there is a general lack of revocation controls in social-networking, e-commerce or indeed almost any cyberspace applications.

An effective system that manages control of data must provide mechanisms enabling the data subjects to truly decide and understand in what they give consent to and allow them to change their mind in the future [96]. The need for such a system is also acknowledged by the White House, in an attempt to protect consumers' privacy and boost innovation in digital economy [263]. In a report published on February 2012, they define individual control and note that "consumers have a right to exercise control over what personal data companies collect from them and how they use it. Companies should provide consumers appropriate control over the personal data that consumers share with others and over how companies collect, use, or disclose personal data" [263].

1.1 The vicious circle

In order to address individuals' privacy concerns in the online environment, one has to understand how the problem occurs. One of the earliest attempts to conceptualise privacy took place in England and the rational that underpinned it can be summarised in the phrase that "every man's home is his castle". In 1763, William Pitt noted that "the poorest man may in his cottage bid defiance to all the force of the Crown. It may be frail; its roof may shake; the wind may blow through it; the storms may enter; the rain may enter - but the King of England cannot enter; all his forces dare not cross the threshold of the ruined tenement" [in [125]]. In his statement, privacy is framed as protection from invasion and it is the right of every

man, even for the most under-privileged people. Its importance raises from the fact that not even the King of England, the highest authority of that time, could defy the right to privacy.

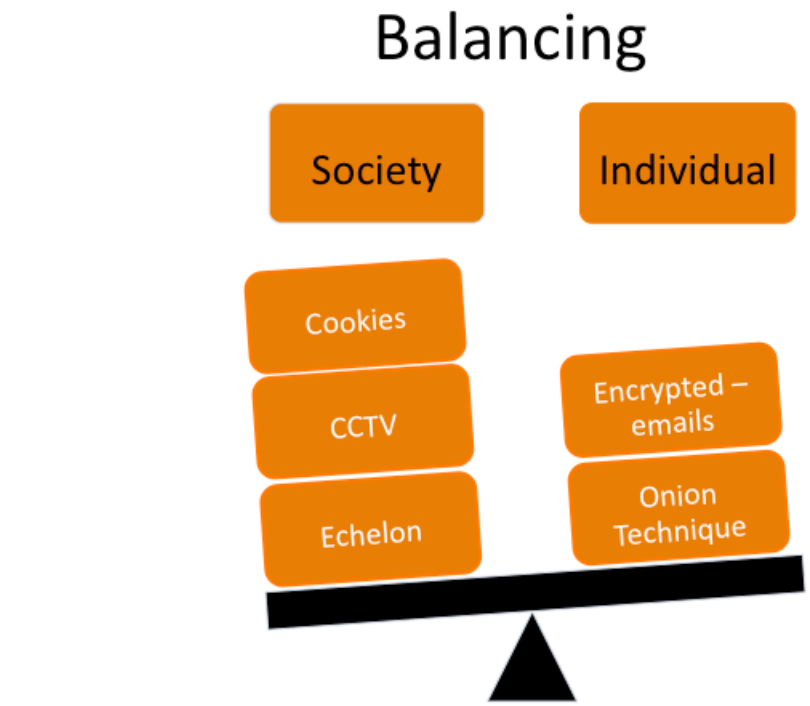
If we subdue the fact that Pitt's definition remains oblivious to women, the idea of privacy as invasion seemed to be effective for more than one hundred years. Until privacy concerns re-appeared as a fundamental factor in the late 1800s in United States, when Louis Brandeis and Samuel Warren argued that new privacy concerns have emerged and novel approaches to privacy, such as their proposition of the "right to be let alone", must be embraced to address them [45]. However, the circumstances that stimulated their position, reveal the root of the privacy problem. As Brandeis' biographer writes [in [40]]:

> Warren had married Miss Mabel Bayard, daughter of Senator Thomas Francis Bayard. They set up housekeeping in Boston's exclusive Back Bay section and began to entertain elaborately. The Saturday Evening Gazzette, which specialised in blue blood items, naturally reported their activities in lurid detail. This annoyed Warren who took the matter up with Brandeis.

As Warren was a prominent and wealthy member of the "higher society", it was natural that the marriage would attract the interest of the tabloids. However, what was unexpected was what Brandeis and Warren describe in their article as "instantaneous photographs" that when published in the newspapers "invade the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that 'what is whispered in the closet shall be proclaimed from the house tops' " [45]. However, according to the journalists, the claim of invasion could not have been justified in a court as the photographer never entered the property. Instead he took the photo while standing on a public place. The authors of the article warned that "recent inventions and business methods, call attention to the next step which must be taken for the protection of the person" [45].

The same consequences that led Warren and Brandeis to publish their article still exist today. Consider the war against illegal drugs in the US: It was thought that using heat sensors to find marijuana growing operations would be acceptable, but in 2001 [Kyllo v United States (533 U.S. 27)] it was ruled that using thermal imaging devices that can reveal previously unknown information without a warrant, does indeed constitute a violation of privacy. With the outburst of technological developments in the recent years, privacy concerns have increased significantly since

new ways of gathering personal information emerged. This affects the ways in which privacy may be either protected or violated, depending on the purpose for which these advances are applied. Examples of technological innovations that have either aggravated or mitigated privacy concerns are illustrated in Figure 1.1.



1

Figure 1.1: Technological developments that favour either society or individuals

The example of Warren and Brandeis demonstrates that legislation and regulatory procedures endeavour to establish functions which seek a balance between individuals' right to privacy and the common good. But every time a balance is found, the use of new technologies alters old norms either in favour of the individual or in the interest of the common good, and new norms and functions need to be re-established to restore the balance, thus forming a vicious circle.

Technological developments always proceed faster than the establishment of legislation and regulatory policies, thus fuelling the vicious circle presented in Figure 1.2 and leading to new privacy concerns. Society is continuously attempting to achieve a balance between privacy and security, individualism and the common good, without ever fulfilling this goal. As Rule argues, privacy is a compromise depicting the

prevailing societal norms, the technological advances and the seek for balance between the private and public sphere [223]. It is these compromises that question the real value of privacy and whether this right is worth protecting [236].

Figure 1.2: The constant development of data collection, aggregation and processing technologies results in a vicious circle as society attempts to seek a balance of protection between the individual's privacy and security of society.

1.2 Establishing the importance of privacy

How can we decide if privacy is important? Why is it worth protecting? These questions have concerned society throughout history. The development of technologies kept the debate around the value of privacy for centuries, but the proliferation of novel information systems strengthened arguments both for privacy advocates and privacy critics from various disciplines such as philosophy, law and economics, rendering the "privacy erupt into a frontline issue around the world" [236].

Privacy advocates argue that privacy is evaporating and soon we will have a sense of nostalgia if we do not act, whereas, critics are based on the privacy paradox, where people declaim their privacy concerns but their actions are towards the opposite directions. Thus, critics claim that there is a general paranoia and people cater for their privacy in abstract [236], while balancing privacy against the common good results in many conflicting interests. In order to ascribe a value to privacy, one needs to understand the benefits and conflicts it causes to society.

1.2.1 Critique of privacy

Critiques of privacy as being socially detrimental, arise from several sources that focus on different aspects of privacy. The most prominent is the communitarian critique which, to quote Etzioni, perceives privacy as a "societal licence [...] that tramps all other rights or concerns for the common good" [86]. Communitarian scholars argue that there is not an absolute right to privacy because it exempts people from obligations of social life [86; 119], associating privacy with individualistic benefits which antagonise the public sphere. However, as many scholars argue, this view is erroneous [236; 124]. The assumption that individual needs and societal interests are always in conflict, results in a fallacious quest for balancing privacy with community needs. As Solove argues, "individualism becomes not an element valued for its contributions to the common good, but a counter-value that stands in opposition to common good", underestimating the contributions of privacy to society [236].

Another criticism to privacy stems from the classical feminists, as they contend that privacy is a right favouring men by concealing the abuse of women at home. Historically, they argue, societal norms have been dominated by men and as a result every husband is the master of the house, a private sphere in which authorities cannot intervene if, for example, a woman is subject to domestic violence [166; 236]. Thus, privacy is the means to protect men and shift the authoritative relations between the two sexes in favour of men. MacKinnon, by paraphrasing Warren and Brandeis view of privacy, claims that "the right of privacy is a right of men 'to be let alone' to oppress women" [166]. This view is valid if we consider common practices in precedent centuries; courts in the 19^{th} century "turned a blind eye to issues of domestic abuse in the name of family privacy" [236]. However, society's power relations between the two sexes have changed and women have strengthen their position. Recent legislated laws are not oblivious to domestic violence and privacy can be used to support the prosecution of offenders as it offers anonymity in cases of rape and violence or forbids any disclosure of data to the public [124].

There is a stream of literature arguing that privacy hinders the disclosure of information critical to establish trust relations and judge individuals' reputations, thus leading to the sharing of discredible information and fraud [211; 85]. Scholars in favour of this argument, emphasise the aspect of self-determination, defined as the ability of individuals to manage information about themselves, claiming that it can be used in dishonest ways to favour the withholding of information or the

dissemination of falsified data. Posner ascribes an economic aspect to this criticism by arguing that erroneous information can prohibit the commercial profitability of businesses. The non-disclosure of vital information tampers with the "economic rationality"of the free market which forces enterprises to depend on assessing vast amounts of information and make rapid decisions [211]. As a result, privacy Cate argues, "may reduce productivity and lead to higher prices" [54]. However, the critic suffers from misconceptions, since privacy can enhance, rather than reduce, the disclosure of data. Individuals feel more confident and "disclose information with brevity" [124] when they have the sense of controlling the flow of their information. Regarding the economic aspect of the argument, Acquisti [3] demonstrated the existence of an economically "negative and statistically significant impact" of data breaches. The impact of the privacy breach is profound in large companies because their reputation is at risk. Hence, the adoption of a system which is able to support the management of personal data by data subjects, can not only enhance the trust relation between customers and enterprises, but it can also boost the disclosure of valid data, offering competitive advantage to those companies that cater for customers' privacy needs.

Questions also have been raised from advocates of civil liberties and human rights who dispute privacy as a right [274]. The rational is similar with the communitarian critique and scholars supporting this argument believe that privacy undermines other primary rights which protect civil liberties. Freedom of speech, liberty and the right to be free from injury are some of the rights which are deemed antagonistic to privacy. To quote Postrel, privacy hinders "the freedom to gather and disseminate truthful information, otherwise known as freedom of speech and the press" [in [236]].This argument, however, assumes that privacy is in conflict with these rights. In a counterexample, someone could argue that privacy can be symbiotic instead. Several authors deem privacy to be synonymous with liberty [101], while focusing on the freedom of speech there have been several journalists who under the umbrella of anonymity or pseudonimity have expressed their opinions freely. The symbiosis of privacy and freedom of speech or liberty ensures that people will be able to express themselves freely, while retaining their dignity at the same time.

Further objections to privacy occur from scholars contending that social surveillance and national security is hampered by the concealment of malicious activities under the protection of privacy. As Pulitzer denotes, "there is no crime, there is not a dodge, there is not a trick, there is not a swindle, there is not a vice which does not live by secrecy" [in [94]]. The argument has gained in popularity after

the events of 9/11 in US, as privacy is deemed to hinder government responses to threats and conceal emerging dangers from terrorism [243; 212]. Posner declares that nowadays "the government has a compelling need to gather [...] vast quantities of information, much of it personal" [212]. In a similar manner, other scholars wonder that "if you have nothing to hide, then what do you have to fear", or "if you aren't doing anything wrong, then what do you have to hide?" [235]. The key problem with this argument is that "it myopically views privacy as a form of concealment or secrecy" [235]. Privacy may enhance openness and transparency in the community since it has a positive effect on individuals' trust relations. In addition , the existence of privacy rights ensure that the governmental surveillance will be conducted with respect to human autonomy and dignity. The weakness of the argument emerges from its assumption that governmental authorities always behave in a benign fashion. However, reasonable reasons to use surveillance may quickly escalate and result in power abuse that causes "intimidation, embarrassment or distraction" [124] as described in Orwell's book titled "1984" [198]. Furthermore, the problem that occurs is not only "Orwellian but Kafkaesque" as well [235]. In Kafka's book "The Trial" [138], the problem raised by the author is the situation where individuals are powerless and vulnerable against the bureaucratic governmental system. The misuse of personal data with the exclusion of data subjects from any governmental process results in a lack of transparency, misjudgement, errors, authoritative abuses [235] and at the end of the book, the condemning of an innocent man, Mr K., to death [138]. Thus, the argument of "nothing to hide" collapses because avoiding behaving in an illicit manner is not enough to ensure individuals' innocence, since, to quote Kafka [138]:

> "- 'One does not have to believe everything is true, one only has to believe what is necessary.'
> - 'Depressing though, said K.'[...] The simple story had become perplexing."

1.2.2 The value of privacy

The vast majority of literature advocates the importance of privacy as essential to an open and democratic society [124; 82; 45; 41; 101; 262]. Many scholars perceive privacy as a right that protects solely individuality, while excluding people from social interaction [82; 45; 41; 101]. To quote Emerson "[privacy] is based upon premises of individualism, that the society exists to promote the worth and the

dignity of the individual.[...]The right of privacy[...]is essentially the right not to participate in the collective life the right to shut out the community" [82]. The value of protecting privacy stems from the self-development and the respect for autonomy. According to Westin "people respite from the whirlpool of social life [...] achieve goals of self-realisation" [262] and thus, develop a "firmer, better contracted position in the opposition of the dominant" [236].

Although the importance of privacy is significant when emphasising on the individual, perceiving privacy in nonconformist terms provokes conflicts with other primary rights and interests and provides solid ground for criticism. When privacy is excluded from social life, inherently it is balanced against common good, resulting in privacy being undervalued. Solove contends that privacy has a social value and its importance also emerges from the benefits it confers upon society [236]. Loss of privacy has a significant impact upon "freedom, creativity, culture and transparency in social life", effects that are not assessed in the equation if privacy is considered as a "stand-alone" right [262]. Since we are not isolated from other people and social interaction is part of our daily lives, we cannot afford to consider privacy as something we abandon when we join social life. Thus, the "value of privacy should be assessed on the basis of its contribution to society" and the value of individualism should be integrated to the social benefit [236], since "we cannot separate the idea of ourselves and our own good from the idea of others and their good" [44].

Privacy offers a wide range of protections to a plurality of issues and as a consequence, its value varies depending on which issue is being addressed. The value that stems from the involved activities should be weighed against the value of contrasting interests and privacy should prevail when the result is best for the society. As Solove argues, "we live in an age of balancing and the prevailing view is that most rights and interests are not absolute and privacy should be reconciled with other interests" [236]. Balancing is a useful approach to resolve conflicts but the critical step to maximise the utility in both parts of equation is to adopt a systematic process that "will be as rigorous and thoughtful as possible" [236].

Privacy is a powerful tool able to shape social norms, alter power relations, ensure democracy and freedom and establish a transparent and open society. Values of the highest importance, but very difficult to be quantified and calculated in mathematical equations. A solution to address privacy concerns while respecting the common good must be sought. Since social trends are shaped by the participation of individuals and the way they communicate, information systems that enable users to manage their personal data can become a barometer for establishing balance.

The research that will be presented in this thesis, has been conducted bearing in mind that privacy cannot be isolated from social good and the balance in the vicious circle must be researched in a manner that maximises both individual and public good. In alignment with Solove's view, I believe that offering mechanisms to the data subjects to control their data is the rigorous process to ensure, even-though the vicious circle will remain, the means by which the exposure to risk can be mitigated more effectively and the balance can be achieved more efficiently.

1.3 Management of controls

The research undertaken in this thesis offers to data subjects, for the first time, the opportunity to express their consent and revocation preferences regarding the collection, usage and further dissemination of their personal data and convey their own privacy expectations. Offering controls to data subjects can create a positive effect on their predisposition to disclose data, thus enhancing transparency and openness in the relationship between the customer and the enterprise. On the other hand, the controls available to data subjects can be designed with respect to the common good, ensuring the social control and the resolution of any conflicts with other interests.

Capturing the process of giving and revoking consent provides solid ground to express privacy controls. Legislation, and more specifically data protection, already requires from the data controllers the acquisition of data subjects' consent before collecting any personal data. In addition, the process of giving consent has been common practice for many years in several disciplines, such as medicine, where privacy is of paramount importance.

This thesis seeks to develop methods by which balance can be achieved via consent and revocation controls over the storage, use and sharing of personal data. The issues addressed arise as a consequence of the disclosure of personal data. Enterprises store personal data about their customers, often not only contact details but various consumer preferences; such information enables an enterprise to tailor its products and services to customer needs. Companies share, buy and sell customer data with the aims of increasing their customer base and obtaining useful marketing statistics. There are even firms whose entire business is to manage and market large databases of personal data about individuals. The central problem is that an individual for whom personal data is held can barely, if at all, control (view, modify,

remove) the storage, aggregation and flow of this data.

While it is necessary, by law, to obtain an individual's consent for the collection of data regarding her or his person, the possible semantic interpretations of such consent can vary widely, possibly opening the way for data misuse and abuse [1; 88; 2]; this can arise when the data controller interprets the given consent in a way that conflicts with the individual's preferences. Furthermore, it is common practice for an enterprise to request 'blanket consent,' which is consent for data to be collected, used and disseminated in any way the data controller deems appropriate; by giving blanket consent, an individual completely relinquishes control over such data. When more fine-grained consent is required, the request is usually accompanied by lengthy details of the collector's privacy policy; research has suggested [65] that individuals rarely study such policies carefully, and this may cause them to give blanket consent to save time and effort.

1.4 Research questions and main contributions of the thesis

Having described the problem that this thesis seeks to address and having established the value of privacy to society that attributes to the importance of the research to be presented in the following chapters, this thesis contributes to knowledge in multiple layers and its novelty concerns different aspects.

Adopting a user-centric point of view, the first research question that I seek to answer is:

> What controls could a data subject perform in order to express privacy preferences and requirements on the management of their personal data?

In order to answer this question, the methodology followed included a multidisciplinary literature study on online privacy. The findings informed the first framework of effective controls and suggested that consent is the appropriate means to offer these controls to the data subject. However, there is a significant gap in the literature, since the functionality of the online environment raises considerable threats to privacy and requires the ability to revoke the initial consent and delete data. As Solove declaims "details about your private life on the Internet can become permanent digital baggage" [238]. However, so far literature perceives consent as an one-off event while the possibility of revoking this consent is not considered.

The second research question which complemented the question presented above in order to provide an appropriate answer is:

What does revoking consent mean?

The approach followed to provide a model for revocation embraced elements from the social sciences discipline. The methodological framework consisted of focus groups and the generated data was analysed with the content analysis approach. The result was a novel conceptual model of consent and revocation, which is the first contribution of the thesis. In addition, different types of revocation have been defined and the term of "informed revocation" has been introduced, extending the existing literature on online privacy.

The next step required the formalisation of the novel conceptual model that can validate a system which offers data subjects the opportunity to express consent and revocation expectations. The research question to be answered in this quest is:

How can consent and revocation controls be formalised?

I conducted a literature survey on the application of formal methods whose aim is to resolve privacy issues. The findings of this review concluded that there exists a gap in the literature, since the dominant approaches consist of languages which address security issues and embrace extensions for catering privacy needs. Privacy is not usually the primary concern of these efforts. Any existing methodologies addressing solely privacy concerns focus mainly on converting personal data to non-identifiable forms and not on controlling the flow of personal data through its lifetime. The difference in the two approaches lays on the sharing of data. In the first approach, the aim is to hamper the disclosure of certain information, whereas the latter approach surveillances any dissemination of data throughout its life-cycle by posing controls. In addition, all the formal models remain oblivious to consent and revocation concepts. A Hoare-style logic was developed, capable to reason and capture the effects of user's consent and revocation expectations in a system. This novel logic is the second and main contribution of this thesis. In order to validate the logic, the use of Maude, a rewrite tool for logic, was deemed necessary to prove that the logic is designed in such a manner that certain healthiness conditions hold. In addition, the logic was applied to formalise the consent and revocation requirements of three different case studies. The pilot Employee case study, revealed ambiguities that were addressed by enriching the logic. The refined and final version of the logic is presented in this thesis. The refinement process and the lessons learnt from the first attempt are described in Section 4.5.

The last contribution of this thesis is a testing strategy, based on the Hoare style

logic, generating tests for systems that handle consent and revocation controls. The research question answered is:

> How might a testing strategy be formed based on formal methods, for a system that offers consent and revocation controls to ensure correct behaviour while services and technological infrastructure evolve?

The result of this effort is a novel and rigorous testing strategy based on formal methods, able to generate testing suites in an automatic way. The approach is technological agnostic and it is validated by creating tests for the first case study. Further validation required the development of a pseudo-code that automatically generates tests. The pseudo-code was adjusted to the needs of the EnCoRe project [83], a project that funded this research for the last three years. The proposed pseudo-code was implemented by HW Communications Ltd *, one of the partners participating in the EnCoRe project who were responsible for the implementation of the system. The novel approach was used to produce tests and reveal faults in the design of the system.

Parts of the research presented in this thesis were presented in conferences. Published papers [10; 11; 201; 7; 8] were co-authored with my supervisors, Professor Sadie Creese and Professor Michael Goldsmith, to whom I am most grateful for their guidance and support throughout this journey. Their guidance ensured that I would not deviate from the research questions and that the methodologies followed would be the most appropriate to provide valid answers, while giving me the liberty to freely develop and express my own understanding of the problem and propose and develop my own solutions. Furthermore, the research presented in this thesis was used to underpin the design of the EnCoRe system and is included in the final deliverables of the project [9; 6].

1.5 About EnCoRe

The EnCoRe project ran from June 2008 to April 2012 and was a multi-disciplinary research project which aimed to "develop technology-enabled solutions for delivering easy-to-use, practical and scalable consent and revocation controls over the use of personal data in cyberspace; and to render these control mechanisms as simple

*The link to their website is http://cyber.hwcomms.com/cyber/

to use as a kitchen tap" [83]. The project was partially funded by the Technology Strategy Board (TP/12/NS/P0501A), the Engineering and Physical Sciences Research Council and the Economic and Social Research Council (EP/G002541/1).

Currently, organisations' procedures for managing personal data may possibly exploit the trust that data subjects place in them when giving their initial consent and disclosing their data. There is not a system in place that can provide controls to the users to express themselves and capture the giving of consent or the dynamic notion of this process required to address data subjects privacy concerns.

In the EnCoRe project we perceived "controls" as the means which enable people to manage the flow of their personal data, by expressing consent and revocation preferences that can be implemented through non-interference and privacy policies. The overall vision of the project was "to make giving consent as reliable and easy as turning on a tap and revoking that consent as reliable and easy as turning it off again" [83]. To this end, EnCoRe project took into account a variety of perspectives, including social, legal, regulatory and technological aspects.

The EnCoRe project endeavoured to provide solutions that would:

- Enable organisations to adopt scalable, cost effective and robust consent and revocation methods for controlling the usage, storage, location and dissemination of personal data.
- Benefit individuals by providing meaningful, intuitive mechanisms which will allow them to control the use of their personal information held by others.
- Help restore individual confidence in participating in the digital economy and so, in turn, benefit the wider society.

The project partners were:

- Hewlett-Packard Laboratories
- HW Communications Ltd
- London School of Economics and Political Science
- QinetiQ Ltd
- Cyber-security group, Department of Computer Science, University of Oxford (October 2011 - present)

- e-Security Group, Warwick Manufacturing Centre, University of Warwick, Coventry (June 2008 - September 2011)
- Centre for Health, Law and Emerging Technologies (Helex), University of Oxford

The project's website is www.encore-project.info and the tweeter account can be found at www.twitter.com/encore_project

1.6 Thesis structure

The thesis is divided in nine chapters. Elaborating on each chapter:

Chapter 2 presents literature surveys of the online privacy and the concept of consent, a description of the regulations legislated for the data protection and a literature review of the formal methods that endeavour to model privacy. The chapter starts with the main concepts of privacy and illustrates their limitations. Based on the dominant concept which perceives privacy as controls, I examine the available controls which researchers believe that can facilitate the management of personal data by data subjects. The most appropriate mechanism is the concept of consent and I explore the origins of the concept, the relevant theories of consent, their application on our daily lives and their limitations. In addition, I present the appropriate legislative efforts that enhance the concepts of consent and revocation and provide legal ground to their adoption by the data controllers. Since the main purpose of this thesis is to develop a formal model of consent and revocation, I study the relevant formal languages that address privacy issues and present their advantages and limitations.

Chapter 3 presents a novel conceptual model of consent and revocation. I describe the qualitative methodology of focus groups adopted to underpin the design of the model and I explain the content analysis approach used to code and analyse the data generated by the focus groups. The analysis of the transcripts resulted in a novel conceptual model of consent and revocation. More specifically, I describe how the findings extend the theory of informed consent with an additional dimension and the definition of different types of revocation. Furthermore, the term of informed revocation is coined to describe a phenomenon observed in the focus groups.

Chapter 4 illustrates the novel logic of consent and revocation, which is designed based on the novel conceptual model of consent and revocation. It is a Hoare-style logic that captures the requirements of a system capable to manage

consent and revocation controls. The expectations of the data subjects are captured with a set of actions that, when performed in completion, assign rights to the participants of the system, create obligations and bound the handling of data with constricting choices. The logic has been refined after its application to the pilot case study and the process of refinement among with the lessons learnt from the formalisation of the case study and more specifically, how these lessons informed the implementation of the EnCoRe system are also discussed in this chapter.

Chapter 5 describes the pilot case study, namely the Employee Data Scenario and presents the formalisation of employees expectations regarding consent and revocation in their working environment. Several use cases depict different requirements which are elicited with the Goal Notation Structure. The chapter illustrates the application of the refined and final version of the logic and provides evidence that the ambiguities acknowledged in Chapter 4 have been addressed.

Chapter 6 presents the formalisation of requirements for a real case scenario, which due to confidentiality issues it will be called the Biobank. There are multiple use cases formalised which illustrate different consent and revocation expectations which patients may have. The requirements are elicited by analysing transcripts from focus groups, specifically designed to capture privacy concerns of patients and cater for the needs of Biobank's researchers.

Chapter 7 presents the formalisation of requirements regarding the consent and revocation controls offered to the data subjects for the Identity Assurance Programme, a mock system inspired by a real project whose name will not be revealed due to confidentiality issues. The context in this case study involves different power relations from the previous case studies, because there are governmental needs to be addressed. The requirements are elicited by analysing documents and discussions from the research community which was assigned to assist in the design of the project. Not all sets of actions, constraints and choices expressed in the final version of the logic are used in each case study, as the business model and the data subjects' preferences change according to the context. The successful application of the logic to three diverse contexts validates its efficacy and richness of expression.

Chapter 8 explains the need for a testing strategy based on the logic illustrated in Chapter 4 and presents the novel approach, which is technological agnostic and able to generate testing suites for a system that manages consent and revocation controls. The model comprises of two procedures and is validated by being applied to generate tests on the Employee case study. In addition, I developed a pseudocode which was adjusted to the technical architecture of the EnCoRe system. The

pseudo-code was implemented in the first EnCoRe prototype, specifically designed for the Employee case study. The testing strategy is successfully implemented in the EnCoRe system and the results of any faults in the EnCoRe implementation are reported in the chapter.

Chapter 9 summarises the research presented in the thesis. The chapter illustrates the contribution to knowledge that this thesis achieves and proposes directions for future work. These directions could extend the logic further to address issues that are not considered in this approach, such as anonymity and improve the automatic generation of testing suites.

CHAPTER 2

The journey from privacy to consent and revocation

What is privacy? Philosophers, legal theorists and scientists from diverse disciplines have not provided yet a satisfactory solution to that riddle. Philosopher Julie Inness declares that the privacy conundrum is in a "state of chaos" [131], privacy advocate Simon Davies contends that "even after decades of academic interest in the subject, the world's leading experts have been unable to agree on a single definition" [73] and the legal theorist Robert Post doubts if a definite and widely accepted answer will ever be given as "privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all" [213].

What can undoubtedly be answered though, is the importance of the concept of privacy which has been highlighted from privacy advocates to privacy sceptics, as it is the catalyst when referring to diverse rights and values that can only be articulated under the umbrella of privacy. As Charles Sykes [245] notes, "privacy is like oxygen; we really only appreciate it when it's gone". In the published literature, privacy is declared as "essential to democratic societies" and recognised as "the heart of our liberty and beginning of our freedom" [101]. It is considered to be the core element for protecting autonomous life, maintaining human dignity [215] and the remedy to preserving individuality [40]. Many scholars intertwine privacy with our senses of rightness and our moralities [142; 202] and argue that the explanation of the privacy enigma is based on the "feeling", the "instinct" and the familiarity with "common usage of language" which is "riddled with paradoxes"[202]. As Lord Philips of Sudbury said, in an abstract and indirect definition of what privacy is, after opposing the ID-card policy[39] in the United Kingdom, "I instinctively and

quite deeply reject [to vote the policy under consideration]. I can't quite find the language to rationalise the depth of my feeling about this" [264].

Privacy is not a "stand-alone" issue but emerges from social interaction. It depends on the context and the environment where these interactions occur and offers opportunities for social association regarding "political expression and criticism, political choice and freedom from unreasonable police interference; it allows non-political participations in family and religion"[262]. It is this social interaction that differentiates our understanding of privacy according to our experiences and the cultural and political environment in which we act, resulting in our privacy concerns to stem from our subjective perceptions [132]. As Simon Davies [73] argues, "privacy means different things to different cultures. In France, it equates most closely to liberty. In America, it is an inseparable component of individual freedoms particularly freedom from intrusion by federal government. Many European countries interpret privacy as the protection of personal data. Since the days of the huge campaign against the government's proposed ID card in 1987, most Australians view privacy as a measure of state power, while the government views it as a set of strictly defined legal rights".

Due to the multi-dimensional, subjective and context-depended nature of privacy, unsurprisingly its prism has been investigated by adopting different lenses that vary widely. In the next section I present the different conceptions in the vast literature on privacy that seek to distil its core characteristics by defining privacy, as Solove [236] argues,"per genus et differentiam". In essence, researchers search for a"common set of necessary and sufficient elements that single out privacy as unique from other conceptions" [236]. In Section 2.2.1 I review the research conducted regarding privacy concerns in online environments. The researchers draw elements from the philosophical concepts of privacy in order to mitigate online privacy concerns and the dominant concept is the perception of privacy as controls. These controls can be captured with the notions of consent and revocation and in Section 2.3 I illustrate how social sciences understand consent and revocation and the proposed models that lead to the theory of "informed consent". These models are depicted on legislative attempts to protect privacy and in Section 2.4 I present the rationale that dictates contemporary privacy regulations. In Section 2.5 I illustrate the formal methods aiming to provide the mechanics for a privacy-friendly framework able to encompass the research undertaken regarding privacy in social sciences and law.

2.1 Conceptualising privacy

I define the conception of privacy as "an abstract mental picture of what privacy is and makes it unique" [236]. The conception of privacy differs from the daily usage of the word privacy, defined as the diverse ways that we speak of privacy when referring to things [236], and I adopt the definition of conception in order to avoid the multi-use of the word that leads to paradoxes [202]. Literature suggests several conceptions in an attempt to describe the common characteristics of privacy. Every conception encapsulates interesting insights and is criticised for its limitations. There is not a clear distinction between the different concepts presented below and many of their characteristics interleave.

2.1.1 Privacy as the right to be let alone

The first attempt to address the privacy conundrum was Aristotle's [74] distinction between the public "polis" and the domestic or private "oikos" sphere. Centuries later, legal theorists encapsulated the problems emerging from this distinction by legislating rights to privacy. Two of the pioneers were Samuel Warren and Louis Brandeis who argued for "the right to let alone" [45], drawing on an earlier view by Thomas Cooley who suggested "the right of personal immunity" [75].

Brandeis's and Warren's seminal article drew attention to the privacy problem and since then it has been the cornerstone for privacy laws in the United States [236]. The authors distinguish the core characteristic of privacy to be "that of inviolate personality" and underline that the value is "found not in the right to take the profits arising from publication, but in the piece of mind or the relief afforded by the ability to prevent any publication at all" [45].

Brandeis's view flourished when he was given the opportunity to advocate and establish the conception of being let alone as a right after the US Supreme Court in *Olmstead vs United States* decided that telephone wire-tapping does not violate the Fourth Amendment since there is no physical trespass [in [202]]. Brandeis in his dissent argued that "they conferred, as against the government, the right to be let alone - the most comprehensive of rights and the right most valued by citizens" [in [236]]. Several US Supreme Court Justices were in alignment with his view and many distinguish legal scholars endorsed his conception such as Bloustein [41] and Posner [211]. The latter however, limits his perception of privacy to "people's right to be free from unwanted solicitations" [210] only.

The right to be let alone preceded its time and set the foundations for more robust privacy conceptions. However, it is criticised of being too broad failing to enlighten the circumstances under which the individuals shall be let alone. According to Parent [202] there are "innumerable ways of failing to let a person alone" which are irrelevant to privacy [202]. Allen[14] contends that "if privacy simply meant 'being let alone', any form of offensive or harmful conduct directed toward another person could be characterised as a violation of personal privacy". Although highly influential as a conception, it still remains quite vague.

2.1.2 Privacy as limited access to the self

A considerable amount of literature conceptualises privacy as "limited access" to the self [101; 100]. The core characteristic of this theory is individual's will to concealment and acting remotely from the others. As Parent denotes "it has the virtue of separating privacy from liberty"[202]. Thus, it is a refinement of the "right to be let alone" theory as it endeavours to describe under which circumstances people could be let alone.

The first version of this theory was illustrated in Godkin's seminal article [107]. Godkin defines privacy as "the right of every man to keep his affairs to himself, and to decide for himself to what extend they shall be the subject of public observation and discussion" [107]. He argues that every individual should decide to what extend public could obtain "knowledge of his affairs"[108]. This work set the foundation for other scholars to underpin their theories. Gross envisages privacy as "the condition of human life in which acquaintance with a person or with affairs of his life which are personal to him is limited" [114] while Den Haag declares that individuals should request exclusive rights to either "watch, utilize or invade someone's realm" [254]. Bok [42] extends this view and distinguishes physical access from personal information, while Allen introduces the concept of inaccessibility as the "apt application of privacy" [14].

There are though subsequent shortcomings in the limited access to self approach. Parent [202] drawing on Bok's dichotomy argues that violations of "physical proximity" could better be explained by other concepts such as personal property while violations regarding access on personal knowledge are considered "limitations on cognitive access that do not imply privacy" [202]. In addition, Solove poses the question of how the grey area between absolute restriction of access and the total accessibility is defined [236]. He suggests that limited accessibility fails to recognise

what violations of access constitute privacy breaches, thus rendering the conception too vague.

Gavison [101] attempts to fill the gap in this theory and as a reply to the criticism she focuses on defining a "neutral" but "coherent concept of privacy" elaborating on what is considered as limited access. Her model consists of three distinctive elements namely "secrecy, anonymity and solitude"[101]. Posner focuses on secrecy and argues for the individual's right to "conceal discreditable facts about himself" [209]. His view is expressed under the prism of economics and by focusing on the profits that such a concealment may offer, he suggests that people "want more power to conceal information about themselves that others might use to their disadvantage"[209]. Jourard adds that privacy should have a dynamic and continuous effect as the "outcome of a person's wish to withhold from others certain knowledge as to his past and present experience and his intentions for the future" [137]. A more refined version of secrecy is expressed by Etzioni that proposes a selective form of secrecy and describes privacy as "the realm in which an actor can legitimately act without disclosure and accountability to others" [86].

Although these refinements of Godkin's seminal theory endeavoured to address the criticism and provide a solid and satisfactory conception of privacy, in their attempt to become more descriptive they ended up in being too narrow. Secrecy and solitude involve only the concealment of the information and once personal facts are publicly disclosed or leaked then according to the theory there can exist no privacy [236]. However people may willingly share their personal facts with a group of people while keeping the same facts private from others [40]. Even in the selective secrecy perception, privacy should not be narrowed in refraining from disclosure but other issues, such as using personal facts for the desirable purpose, should be taken into account [236]. Furthermore, there are situations where there is no "reasonable expectation of privacy" [236] when people for example act in public places. As Parent suggests limitations on access, no matter how applicable they could be, are not privacy themselves but "safeguards" of what privacy stands for [202].

2.1.3 Privacy as personhood and autonomy

Another stream of literature conceives privacy as the means for protecting autonomy and personhood. Gavison, although a prominent advocate of the limited access theory, recognises the value of privacy to be lying on the principles of autonomy, liberty and freedom [101]. Considering autonomy, privacy could be "the freedom to

decide and to act in public or private as one deems appropriate, without government interference" [15]. An example where autonomy is endangered, as Benn declaims, is when people are under surveillance and in essence their freedom is restricted since the "observed becomes aware of himself as an object having a determinate character" [36]. Riley extends this view by arguing that there should be an "understanding that certain aspects of people's lives are sacrosanct and only shared in cases of justifiable legal requirements" [219].

The term personhood was coined by Freund to identify "those attributes of an individual which are irreducible in his selfhood" [in [66]] and the emphasis is on the integrity of the person. The foundation of the theory is Brandeis's concept of "inviolate personality" [45] and Bloustein builds upon suggesting that privacy is the concept which safeguards people against "conduct that is demeaning to individuality and to personal dignity" [40]. Davies concurs and identifies integrity amongst other elements as a core characteristic of privacy [72].

Criticism of the theory evolves around the association of privacy with liberty and autonomy, arguing that these notions should be independent. However, this view is opposed by DeCew, who believes that these conceptions can interleave [75]. Further criticism highlights the lack of a clear and articulated definition of what personhood is [236]. In addition authors that discuss individuality do not present satisfactory descriptions of what individuality is [114]. Furthermore, the case where personhood or individuality is equated to autonomy leads to a vicious circle as "to call an individual autonomous is simply another way of saying he is moral free and to say that the right to privacy protects freedom adds little to our understanding of doctrine" [222]. This vicious circle is enhanced as defending dignity and personhood against surveillance and intrusion requires the state to establish its own understanding of the conception, hence restricting the autonomy of the individuals to enforce themselves what is appropriate according to their beliefs. In addition, there might be situations where loss of dignity does not imply a loss of privacy (for example having to beg in order to survive constitutes a loss of dignity but it is not a privacy breach) [173].

Despite the criticism, the personhood conception could be complementary to other theories or underline important privacy breaches. Beardsley suggests that "the norm of autonomy is what gives our obligation to respect another's right of selective disclosure" [27], strengthening the limited access conception. Another interesting observation by Solove, responding to Rubenfeld's criticism [222], is that although the theory of personhood and dignity is not robust enough it reveals the

problem of aggregation. Whereas minor and isolated pieces of information may not be considered as privacy breaches, when combined together they could begin to "portrait our identities" [237], rendering personhood a necessary element for every privacy theory.

2.1.4 Privacy as controls

The perception of privacy as a means of information controls about ourselves dominates the literature [33]. Alan Westin set the cornerstone of this theory by defining privacy as "the claim of individuals, groups or institutions to determine for themselves when how and to what extend information about them is communicated to others" [262]. Introna concurs and distinguishes one aspect of privacy to be the "control over personal information" [133], whereas according to Miller, for a privacy theory to be robust it must provide the "individual the ability to control the circulation of information relative to him" [176]. Fried extends this view by arguing that the absence of information "about us in others minds" does not necessarily constitute privacy but it is the "control we have over that information" that matters, perceiving controls as a subset of the limited access theory [97].

One of the limitations of this theory is that it neglects aspects that are not related to information, rendering the theory narrow [236]. Furthermore, not every loss of control implies loss of privacy and vice-versa [75]. Schoeman imagines a situation where a person is isolated in a deserted island; he practically has no control over who possesses information about him however there is no privacy breach, rather "his problem is that he has too much privacy" [227]! Another question that seeks a clear and solid answer is what information the individual will exercise control on. Schoeman argues that "to determine what information could be available to others presumes privacy is something to be protected at the discretion of the individual to whom the information relates"[227], raising issues regarding the moral aspect of privacy as it is not only a single person's prerogative but it should invoke society decisions as well. But even if we manage to achieve the perfect balance and the individuals are able to draw the limits of what information they can control in a satisfactory way, there is still the question of whether the individuals are in the position to act in a meaningful or informed manner. As Schwartz notes, "there are disparities in knowledge and power in bargaining over the transfer of information" [228].

Parker attempts to reply to some of the criticism by defining the concept of

personal information upon which individuals could impose controls. He suggests "control over who can see us, hear us, touch us, smell us, and taste us, in sum control over who can sense us is the core of privacy" [205]. However, his definition is very broad. Murphy's attempt shares the same fate when he describes personal information as "any data about an individual that is identifiable to that individual"[188]. Identifiable information should not always be considered as private and a counter-example for such a definition is that of a famous actor where people are aware of his profession but it is not considered private information.

Acknowledging the problem, a stream of literature endeavours to limit the scope of personal information by introducing the element of intimacy and reallocates the value of privacy from posing controls to personal information in general, to posing controls to the development of relationships. Fried defines privacy as "control over knowledge" essential in enhancing "fundamental relations of respect, love, friendship and trust" [97]. He concludes that "intimacy is the sharing of information about one's actions, beliefs or emotions which one has the right not to share with anyone", hence circumventing the problem of morality. Rachels concurs and recognises that the value of privacy lies "between our ability to control who has access to us and to information about us, and our ability to create and maintain different sorts of social relationships" [218]. Inness builds on that and specifies that the value of intimacy emerges from "the agent's love caring and liking" and the control of a person on matters like that should "cover choices on access to herself, the dissemination of information about herself and her actions"[131].

Another challenge for the control of information theory is that theorists do not elaborate on what controls could the individuals exercise. One solution proposed by Westin is to understand controls as a form of ownership. He argues that "personal information, thought of as the right of decision over one's private personality should be defined as a property right" [262]. Locke contends that "every man has a property in his own person" an argument that underpins the intellectual-property law [161]. However, there are serious limitations when treating controls as intellectual property, since information may be created in a group of people or can be derived from an original piece of information, posing dilemmas on who the owner of the information is.

2.2 Privacy in the online environment

According to Solove, in a volatile environment where people depend upon services and benefits which are delivered via the Internet, privacy is evolving [236]. While individuals seek more refined mechanisms to control their information, from innovative technologies to sophisticated business practises and legal requirements, the conceptualisations that attempt to address the privacy conundrum narrow the understanding of the problem and our ability to provide effective solutions [265; 55].

Solove believes that in order to capture the "dynamic and evolving" nature of privacy in an Internet-based society, we need to distance ourselves from describing general and abstract elements in request for a "common denominator" and focus on viewing privacy with many different lenses, investigating different contexts and defining problems instead of offering solutions. He underpins his privacy theory on Wittgenstein's concept of family resemblances. Wittgenstein, in his attempt to solve problems on logic and language, developed his theory of activities that although they have "no one thing in common" somehow "are related to one another in many different ways" [271]. Hence, instead of identifying the common element the aim is to draw the common canvas of elements and at the end there is not a single answer but a "variety of answers depending on a variety of factors" [102].

Solove emphasises on privacy problems and creates a taxonomy for privacy by identifying four basic groups of activities that lead to problematic situations. These are:

- Information Collection
- Information Processing
- Information Dissemination
- Invasion

He analyses those activities further by identifying specific privacy problems. The most important of these problems are: *aggregation* which he defines as the "combination of various pieces of data about a person"; *identification* which is "linking information to particular individuals"; *insecurity* which depicts the situation where stored information is poorly protected; *secondary use* defined as the process of using collected information for different purposes than those initially collected; *exclusion* defined as the failure of an individual to be aware of data that others possess about

him and become part of the handling the data; and increased accessibility as "amplifying the accessibility of information" [234]. These are all problematic situations that individuals seek to address with the appropriate controls.

Another interesting conception that captures the evolving element of privacy is Nissenbaum's theory of "contextual integrity" [193]. She suggests that processing and gathering of information may be allowed in some context as long as the process "obeys the governing norms of distribution with in it", suggesting some kind of control, but prohibited in others [193]. For example, a doctor may discuss health issues with the patients or other colleagues but should refrain discussing these issues with their friends. Both Nissenbaum and Solove attach to privacy a dynamic notion but they do not quest the controls that can be appropriate for the individuals to adequately address the privacy issue in an online environment. It is this question that empirical researchers endeavour to answer.

2.2.1 The notion of controls in online privacy

In the literature on online privacy, several surveys have identified as a primary consumer's concern the way data is being collected and processed by private and public companies via the Internet [181; 151; 200; 232]. More specifically, according to Harris, 78% of online consumers would disclose more data if privacy practices were transparent [261], as online privacy is "a salient issue for users" of the Internet and they prefer websites that they believe are "privacy protective" [250]. Lach estimates that 71% of respondents request stricter legislation for online privacy [157] and Clarke [61] argues that these concerns are the threshold for a new debate about trust in online communications. He estimates that building a trusting environment will amplify consumers' concerns regarding privacy [61]. However, these studies do not explore the nature of those concerns. Asking "how concerned are you about threats to your personal privacy in America today" [Equifax, 1990 in [179]] may shed some light on the importance of these concerns but it will not provide valid data for the nature of those concerns [233].

Once the importance of privacy in online transactions had been highlighted [181; 151; 200; 232; 250; 61], researchers started to explore the nature of privacy concerns by focusing on the collection, use and the dissemination of personal information online. Control issues are dominant in these surveys as, according to Alge [13], a significant factor regarding privacy concerns is "how much control one believes he or she has over handling of information (use and dissemination)" and researchers

are drawing elements both from the "privacy as controls" conception [262] and from Solove's theory [236]. However, identifying and proposing specific controls is a delicate issue since "we have little idea of the ways in which people in their ordinary lives conceive of privacy and their reactions to the collection and use of personal information " [118].

Smith et al. developed and validated an instrument for measuring individuals' privacy concerns regarding organisational informational practices [233]. Based on the literature on privacy conceptions, they underpin their instrument on four central dimensions that affect individuals' concerns, namely collection of personal information; unauthorised secondary use; errors in personal information; and improper access to personal information. They verified the correctness of their instrument by conducting a series of interviews and surveys where as they reported, it became clear that participants were unaware of the processes that companies follow to collect data [233]. Culnan and Armstrong support Smith's findings and declaim that "privacy concerns can be mitigated through fair information procedures" [68]. In addition Sheehan argues that online users should be better informed of their "rights and responsibilities" [230]. The findings in these researches concur and verify the Fair Information Practices framework released by the OECD [95].

Malhotra et al. adopt the instrument proposed by Smith and their colleagues and extend the four dimensions of the privacy concerns by conducting an online marketing survey where they consider whether an individual has "control" over the data. Building upon the work of Smith et al., they distinguish behaviours towards the "control over the use of personal data, awareness of privacy practices and how personal information is used" [242]. From a managers' perspective they suggest that it is important to ensure that the correctness, collection and use of personal information, both inside and outside the borders of the company, could be easily verified by the users. From a customer's perspective they highlight the importance of controls and they indicate some limited options such as "controls to add, delete and modify information at will" [167].

Further qualitative studies verify the four pillars that influence consumers' privacy concerns and provide refined models by examining the interrelation of controls with each dimension separately. A stream of surveys focuses on the collection of data and investigates the dual relation of peoples' perceptions towards a privacy-friendly data-collection process, with their intention of disclosing data and their "refusal and misrepresentation" of that data [239]. Findings suggest that when consumers were ensured of fair information practices, they were willing to divulge

more personal information [68; 77]. In the case where a lack in transparent privacy policies is evident, Hui et al. contend that consumers as a counteraction tend to disclose falsified information or decrease the amount of data they disclose [127].

Another branch of quantitative studies rather than revolving around the amount of information released during the collection process, examine the type of information that individuals disclose. A survey conducted by Son et al. concluded that consumers provide information that they believe does not jeopardise their anonymity [239]. Furthermore, their findings suggest that some types of data are more sensitive, hence more hesitantly disclosed than others, while they also investigate whether firms could derive further information from sensitive data in an unsolicited manner [239].

Other studies consider the tensions in the handling of data, identifying usage as the protagonist factor in consumers' privacy concerns. As Hann notes "among the concern dimensions, we find that the respondents value improper access and secondary use to be more important than possible errors [in collection]" [115]. However, Brown's et al. survey provides contradictory findings as consumers' concerns focus not only on the unauthorized secondary use but also on errors in personal information [46]. Cranor et al. concur that the purpose for which information has been disclosed is a significant factor and contend that any violation may result in consumers' dissatisfaction [65]. Culnan et al. report that when the collected information for a specific purpose is subsequently used for a different reason it should be perceived as a privacy breach and they introduce the notion of consent to control whether the secondary use occurs "without the knowledge of the consumer" [67]. They note that consumers should control secondary use of their personal information by "objecting to other uses when information is collected for one purpose and used for others" [67].

An interesting study by Van Dyke et al. verifies that attention should be drawn on how companies misuse the data of the users and provides a refined model of controls by introducing the processing principles of notice, choice and access. They coin the term of "construct privacy empowerment" which they define as a "psychological construct related to the individual's perception of the extent to which they can control the distribution and use of their personally identifying information" [255]. Hoffman et al. concur and argue that controls on usage of data could tamper the power relations between the individual and the companies and "shift the balance of power from service providers, who have traditionally held power, to the consumers who have traditionally been powerless" [123].

Awad et al. focus particularly on one aspect of Van Dyke's model, and examine the personalised services that some enterprises offer [23]. Further findings suggest that companies which provide online services, should be very cautious when they send their costumers targeted marketing messages or attempt any other form of unsolicited communication [117; 65]. In addition, any further dissemination of data to other companies has subsequent consequences in consumers' privacy concerns [65].

Regarding consumers' access and choice, Cranor et al. contend that when users are having access to their information "within company data bases" their privacy concerns are mitigated. In addition they identify a positive relation with consumers' attitude to disclose more data [65]. In a similar manner, Acquisti et al. suggest that once the purchase is finished and the consumer has no further access to the data it may affect their perception on the secondary use of data, since consumers do not have any indication of how their data is handled [4]. Other studies examined the accessibility of data by users from an economical perspective and concluded that organizations gain "competitive advantage through customer retention" when they facilitate procedures to enhance customers' privacy [68]. Tsai et al. concur and further believe that there may be a niche space for business to profit as "when privacy information is made more salient and accessible, some consumers are willing to pay a premium to purchase from privacy protective websites" [250].

Only a limited number of surveys focus on further dissemination of data from the data controller, where the data subject initially disclosed the data, to another third party. Stewart et al. examine the relation between further sharing of personal data with privacy concerns. They identified that the refusal of the dissemination of data further down the chain is a "small degree of control" and that "consumer control is also communicated on behalf of companies when they state on application forms that any personal information collected will not be shared with any other organisation" [242]. A survey conducted by Cranor et al. declaims that onward sharing has a negative effect on people's trust in companies as "privacy policies do not have [any more] significant value to consumers" [65]. Cranor's research is not the only one suggesting a connection between trust and privacy concerns. Camp et al. contend that when the concerns are "elicited by the merchant's behaviour, the individual may lose trust in the merchant" [51].

Several findings are consistent with Camp et al. [51] and acknowledge that there is a positive relation between privacy concerns and a decrease in trust [160]. Tan et al. developed a conceptual model of trust for an e-commerce environment, where the trust level of any transaction consists of "party trust and control trust" which

are the fundamental "mechanisms that ensure the successful performance of the transaction". It is generally agreed that some level of trust is required in order for people to engage in e-commerce transaction [246]. Dinev et al. [77] verify in their research that limited control over data is associated with trust, which is a fundamental element for the growth of online services [123].

Responding to consumers' needs for establishing a trust relation with online providers, several "trust seal" programs have emerged online to "facilitate the relational exchange of personal information" [164]. An example worth mentioning is TRUSTe, a non-profit, privacy seal program "dedicated to building consumers' trust and confidence on the Internet" [35]. When the logo of the TRUSTe is displayed on the website, the provider declares the processes followed to collect, use and disseminate personal information. TRUSTe, as a third party, certifies that certain functionalities, such as notification for the disclosure practises, protection of the data from misuse and access to consumers to prevent inaccuracies in their information, are available [35]. However, surveys reveal that "trust seal" programs have no effect on the disclosure of personal information [127] while people do not seem to understand privacy seal programs [65].

Both quantitative and qualitative research in the literature provide valuable findings regarding the relation of privacy concerns with controls and they identify specific areas appropriate for control mechanisms. However, they fail to shed more light on the means that can enable these controls. Whitley argues that "the understanding of user-centric control is an impoverished version based on earlier understanding of technology" and that information privacy literature provides an "implicit and limited" view of controls [267].

Implicit as there are a few references in the literature that imply some kind of control meaningful to the consumers mostly envisaged, although not clearly articulated, through the notion of consent. For example, Culnan et al. identify as an effective mechanism the "ability of individuals to remove their names from mailing lists. Milne et al. [179] expand this view by suggesting that people should be able to remove any type of data while Son et al. provide some rather limited controls suggesting that consumers could opt-out "from the database to receive targeted marketing messages" registered initially without their knowledge [239], a view also adopted by Malhotra et al. [167]. A different approach from Son et al., visualises as controls the sharing of negative experience and feedback with other users, aiming at ruining enterprises' reputation [239].

The literature provides a limited perception of controls because it mainly evolves

around the disclosure of data and the understanding of how the data is used, whilst in an "Internet enabled society it is increasingly important to understand what can be done to control this further use and reuse" [265]. An effective mechanism, only implied or briefly mentioned from some researchers [67; 233], to represent these type of sophisticated controls is the *notion of consent* which can provide useful insights to when, how and by whom the personal information is used. In addition, as voices for the "right to be forgotten" [171] become more vocal, the *notion of revocation*, a concept poorly investigated, becomes essential to every theory that targets to address privacy concerns.

2.3 The literature on consent and revocation

The concept of consent has attracted the interest of many researchers in various disciplines. It is a vital requirement for establishing and preserving the ethical element of every discipline as it is "at the heart of codes of research ethics" [269]. Naturally, consent originated and has been extensively studied in the field of medicine and bioethics [135; 269]. The cornerstone of the idea of consent in medicine is the safeguarding of patients' autonomy. The core elements ensuring that individuals' autonomy will not be eroded is the voluntarily nature of consent, since consent should not be given under coercion, and the informed manner under which the patients should act when giving their consent.

In the recent years, as there is overgrowing evidence that providing a user-centric control model in online transactions mitigates privacy concerns, researchers recognise that the element of consent as applied in the field of medicine, can become a basis for understanding and implementing controls [265; 159; 177; 123]. Kerr et al. argue that consent "becomes privacy's linchpin" and envisage it as a "nexus", providing the "interface between human beings and our increasingly automated information gathering systems" [147]. Whitley [265] deems that consent can be an effective and important mechanism when determining how and when the data will be used and disseminated further. In general there is a belief that consent can adequately address the problems and the criticism emerging from the literature on online privacy. In order to comprehend how the notion of consent can facilitate the concept of controls in online transactions, we need to obtain an insight to its origins, the rationale under which it was applied and the limitations of the concept in the field of medicine, the discipline where the idea of consent was initially conceived

and widely applied.

2.3.1 History of consent

There is an "institutional myth" that the idea of consent was conceived as a remedy to the horrifying experiments conducted by doctors that facilitated the operation of German concentration camps during the World War II [122]. Following the end of the war, several trials were conducted by the Americans and amongst those accused for crimes against humanity were the Nazi doctors. The culmination of these trials was the Nuremberg Code that established "the right of a research-subject to give a voluntary consent"and "the right to withdraw from research" [122]. The rationale was to respect the autonomy of the data subjects and protect it against any hazardous research [17]. However, Grodin identifies fragments of consent in the literature that preceded the Nuremberg Code. In 1767, a British court prohibited doctors from trying new instruments on patients without their approval, whereas in early 1900 in Prussia a primitive version of consent was developed. Ironically, this version was later adopted and legislated into law by the Nazi Weimar Republic. Nazi's were the first to prohibit research on animals and individuals but later these laws were neutralised when applied to prisoners of war [113]. The same challenge arose for the Nuremberg Code as American doctors did engage in hazardous experiments as well. It is an indication why the idea of consent was not adopted by doctors in daily practise until the early 1960s since, as Katz, explains obtaining consent "is a good code for barbarians but an unnecessary code for ordinary physician-scientists" [143].

It was not until 1957 and after several trials where doctors were accused by unsatisfied treated patients [221], that the idea of consent was extended by the term of "informed consent". In the case of Salgo v Leland Stanford the court wondered whether the patient was appropriately informed prior to giving his consent for the treatment [159]. The first articulated authoritative definition of informed consent was captured by the Helsinki Declaration in 1964, developed by the World Medical Association, where:

> In medical research involving competent human subjects, each potential subject must be adequately informed of the aims, methods, sources of funding, any possible conflicts of interest, institutional affiliations of the researcher, the anticipated benefits and potential risks of the study and the discomfort it may entail, and any other relevant aspects of the

> study. The potential subject must be informed of the right to refuse to participate in the study or to withdraw consent to participate at any time without reprisal. Special attention should be given to the specific information needs of individual potential subjects as well as to the methods used to deliver the information. After ensuring that the potential subject has understood the information, the physician or another appropriately qualified individual must then seek the potential subject's freely-given informed consent, preferably in writing. If the consent cannot be expressed in writing, the non-written consent must be formally documented and witnessed (World Medical Association, 1964) [in [265]].

This definition emphasises on the autonomy of the patient and the idea of consent is introduced to establish an environment where any violations will be prohibited [28; 89]. The Medical Association dictates two requirements to ensure that the process of informed consent is valid. Firstly, the patient should not be coerced or forced when giving consent and secondly the data subject must act in an informed manner.

Regarding the "freely given" consent, this requirement underlines the situations where the patients are coerced either because there are not being given an alternative or becasue they are being manipulated when the consent choices are engineered. In the latter case, the default choices of consent are extremely important [163]. Due to the aspect of " voluntarily agreement", consent should be specific and propose a detailed aim since "one never agrees in an vacuum; rather one agrees to something or with something" [147]. In addition, consent should be obtained before the disclosure of any data [71].

Since consent precedes the disclosure of data, it is required that patients should be informed of all the benefits and risks regarding what they give consent for, rendering a clear and coherent informing process essential. In addition, the data subjects should also realise the consequences of their refusal to give consent. The issues that emerge relate both with the understanding of the description regarding the purpose of the consent and with the comprehensiveness of the procedure followed to provide information to the patients [175; 144]. Elaborating on patients' comprehension regarding the purpose of consent, Williamson [269] proposes Feinberg's concept of "inducement set", where a subject must be aware of all the true propositions that in relation to his beliefs will determine whether or not to disclose the data [90]. In addition, there are categories of patients where certain circumstances prohibit them from performing in an informed manner.

Faden and Beauchamp, reflecting on the desired conditions for informed consent, developed the most important theory of "informed consent", deriving significant insights on the procedures for obtaining consent [89]. Drawing on concepts from moral philosophy and law, they describe a theory of informed consent based on the principles of respect for autonomy, beneficence or welfare of the data subject and justice or fair treatment. The authors emphasise on informed actions rather than informed persons, since sometimes uninformed persons could act in an informed manner [89]. Further, they provide five core elements that underpin their principles for a valid informed consent. These are:

- The disclosure of relevant information regarding the process prior to data subjects decision.
- Comprehension of the benefits and risks of each treatment and consequences of the denial of giving consent.
- Voluntariness of giving consent.
- Competence of the patient that gives consent.
- Alternative options, where the patient must be provided with the choice to prefer other solutions.

However, from the core elements of the theory of informed consent emerge a two-fold limitation. The first aspect considers the manipulation of the data subject when choices are engineered and the second questions the intellectual capacity of data subjects which under certain conditions prohibits them from acting in an informed manner.

2.3.2 Consent in the literature on information privacy

In the last 30 years, informed consent was adopted and applied in the context of online transactions and it is now considered to be a core feature for mitigating privacy concerns [99]. It is mentioned in data protection regulations and policies, referred in best business practices and linked with fair information processing principles. However, as a concept, it is arguably not completely understood in the online community. Questions remain unanswered regarding how consent can function and more importantly how it can be managed and implemented by the public and private sector [99; 267].

Recently, there have been efforts to provide guidance on how to apply the idea of informed consent in online environments and many researchers recognise the problems and the limitations created from such an endeavour [99]. Researchers adopted Faden's and Beauchamp's theory of informed consent and considered how it can be applied in an online environment where the individuals will be able to control via consent the collection, use and dissemination of their personal information [37; 98; 177; 99]. The most notable example is Friedman et al. approach, where they embrace the core elements of the informed consent theory and suggest a model with one additional element that of agreement. They define agreement as a "reasonably clear opportunity to accept or decline to participate" [99].

However, several issues emerge when the model is applied in online environments that raise questions regarding the core elements of the informed consent theory. Focusing on the disclosure of information, it must be clear to the individual how the information will be collected, who will be able to obtain access to that information, for which purposes the data will be used and under which circumstances the information will be disseminated further. These are controls mentioned in fair information practice principles and a clear articulation and understanding of these practices is of crucial importance [98; 146].

Departing from disclosure, comprehension of procedural practices is essential since the data subjects are required to perform in an informed manner. Towards that direction, they must adequately understand the privacy policies and the fair information practices under which businesses operate. Studies suggest that standardization of these policies and practices has a significant positive effect on the level of understanding [144]. However, as surveys conclude, there are worrying signs that only a limited number of data subjects will read the privacy policies before consenting to them [178; 175].

Furthermore data subjects must be aware of the benefits and the risks undertaken when they agree to give their consent. Unlike medicine, in an online environment a dynamic model for consent is required as the continued use of data by the enterprises may lead to a secondary use of that data. As a consequence, if the initial consent remains static, it is no longer valid. Current business practices obtain data subjects' consent in the appropriately signified time, but unfortunately its static form is considered to be an "indication" that the individuals have provided consent for the use of data under different purposes. Further issues emerge from the secondary use of data as it may enable the deduction of information about data subjects that individuals did not mean to reveal [98]. This could be achieved by

aggregating their data with information derived from other sources [236]. Hence, not all information is of the same importance, as data considered as "sensitive" could reveal further information when processed differently or aggregated with other data. In these cases a detailed and clear description of the benefits acquired when data is disclosed is crucial, as the risk in the event of a privacy breach can have devastating consequences [146].

Voluntariness suggests that consent should be freely given. This element emphasises on situations where data subjects may be manipulated when the consent choices and the default values offered to them are engineered [163; 25]. There is a stream of studies that draw from concepts of psychology and decisional theory and reveal the important role of "choice architects" [147]. These people "quietly skew individual decision making while preserving the illusion of free choice" [147], thus impeaching the premise of the freely-given consent.

Based on psychological barriers, Kerr et al.l declaim that people inherently develop a tendency prohibiting them from freely giving consent and more importantly minimising the possibilities of withdrawing that consent in the future [147]. To corroborate their point they draw upon decisional theory and the concept of "subjective utility" [147]. Subjective utility suggets that individuals' personal value of an outcome correlates with the expected time of that outcome. More specifically the subjective value of a loss or benefit decreases as the time-expectancy of the outcome increases, a suggestion which is in alignment with the prospect theory [162]. For example, individuals will prefer to gain £50 at the present time rather than gaining £100 after 5 years. Furthermore, the decreasing rate of benefits is inverse proportional to that of loses. Thus, gains become less good in the future while loses become less bad [197; 3]. Hence, when an individual gives consent to gain some benefits, the loss of control that might occur in the future is less valued. On the contrary, when they intend to withdraw the initial consent the loss of their benefits is "higher weighted" than the regain of that control in the future [147; 139].

Further issues regarding the voluntariness of consent emerge when individuals are manipulated. Based on normative behaviour, Marx suggests that businesses endeavoured to universalise practices and establish them as normal behaviour coercing individuals to react in a manner that rarely reflects their privacy concerns [168]. For example being asked for your personal details by a cashier tends to become the normal shopping experience. Reinforcing Marx's point, Kerr argues that bounded rationality, which suggests that people have knowledge limitations thus they can make wrong decisions [231], enhances the "misdirection" of the consumers while

preserving a sense of illusionary choice [147].

Another aspect that emerges from the element of freely-given consent, is the problem of "Hobsons choice". Hobson's choice describes a situation where the individual has no alternative but to give consent to a specific action, rendering the process of informed consent meaningless [265]. Absence of choice may occur when data is collected for governmental use, for security reasons or in cases where individuals have to provide personal information to acquire access to services. For example, the popular electronic games EA Sports requires full access on individuals' computers before permitting them to install and play any of their games, providing them with no other choice but not to play the game.

Competence refers to demonstrating the cognitive, mental and physical capabilities required for an individual to give informed consent. It is of paramount importance in online environments since acquiring access on the Internet is easily achieved by almost anyone. The issue is mentioned in the Article 29 Working Party, where certain categories of individuals, such as children and individuals with mental disorders are considered vulnerable online because they lack the capacity to give appropriate consent [216].

Regarding children, their capacity to decide whether to disclose their personal data relies on several factors and the choice until they are able to demonstrate competence is delegated to their parents [79]. In the case of adults lacking capacity, it is essential to acknowledge that it is not a constant situation and a more dynamic response should exist. The responsibility for giving consent on their behalf is delegated to relatives and primary carers [128].

Agreement refers to a clear, legible and visible online opportunity for individuals to either accept or decline giving consent. It should be an on-going concept and individuals should be able to withdraw their consent in the future. There are cases though that not all forms of agreement are explicit, leading to the concept of implied consent where by explicitly giving consent to gain access into a service, the individual has in essence consented to all the activities that might occur in that context. In order for the implied consent to be valid all the other elements of the informed consent must hold. For example giving consent to pay by credit card online, implies that the individual agrees on his bank acquiring the data of the transaction.

Informed consent reveals the potential risk or benefit of an online transaction and provides an opportunity to the data subjects to choose whether to accept or decline to participate. It can be the appropriate mechanism to enable users to control the collection, use and further distribution of their personal data, mitigating their

privacy concerns and enhancing their trust towards data controllers. However, there are limitations that current consent practices fail to address. A model of consent that will provide meaningful control over personal information must embrace the dynamic nature of the online environment and reflect people's behaviour. Acquisti et al. declare that "we need to incorporate more accurate models of users' behaviour into the formulation of both policy and technology" [5]. To-date, understanding the management of consent in information systems is limited and as a consequence, controls applied in online interactions with the form of giving consent remain archaic.

Unfortunately, current practices conceptualise consent as a unique transactional moment, leading to a static and rigid model unable to handle the ongoing and rapidly changing procedures of collection, use and disclosure of personal information. Thus, the current implementation "fails to recognize the unique role that consent is meant to play as the nexus between people and information technology" [147]. A refined model for informed consent, able to address the issues of disclosure, voluntariness and competence, must be based on a dynamic perception that will enable users to change their minds and revoke their consent. Hence designing a model for revocation of consent is crucial for a holistic understanding of the management of informed consent. Furthermore, a dynamic model that offers revocation choices can mitigate the psychological barriers as users can always make the necessary changes to regain the control of their data because it is no longer "an all-or-nothing, take-it-or-leave-it, instantaneous transaction" [147].

2.3.3 Revocation in the literature on information privacy

Although the literature presented above suggests that controls is the dominant concept for mitigating privacy concerns and consent is a well understood and applied concept in medicine, the conceptualisation and management of controls as giving consent in online interactions is limited [265]. Many studies restrain their perception of controls to fair information processing, abilities to access data and opt-out choices, while others interpret consent as a rigid and static process rendering the controls of consent inappropriate for the dynamic online environment. Giving consent to an on-going and volatile procedure of collecting and processing personal data, such as those implemented by business operating online, presupposes that individuals will be offered the opportunity to revoke their initial consent either by replacing it with a different one or by indicating that their consent is no longer valid.

However, the literature that conceptualises control of personal information as

revoking consent is still a stream in infancy. To my knowledge there has not been any proposals or designs for a model of revocation equivalent to that of informed consent. Those researches that suggest revocation mechanisms to enhance the control of personal data, present simplistic choices, such as opt-out mechanisms from targeting management practices [239; 167], the removal of data and the sharing of negative experiences [179].

Kerr et al. recognise the gap in the literature and they suggest that "consent provides the fulcrum for understanding" of controls as means for protecting privacy as long as the notion of "consent as on-going agent" is proposed [147]. Towards this direction Whitley [265] identifies several encounters that could alter individuals choices of consent and argues that revocation of consent could have multiple interpretations. He lays the foundations for a model by suggesting the differentiation between revoking consent for the collection of personal data and revoking consent for processing it. Whitley also acknowledges that further research is required for understanding "what is meant by revocation in the context of organisational systems"[265]. In addition, the recent shift to legislative efforts regarding data protection in the European Union, renders the development of a revocation model more imminent than ever [216].

2.4 Legislating privacy and consent

The evolution of Information and Communications Technologies (ICTs) and the adoption of innovative business models craft the processing of personal information. Thus, legislation regarding the data protection, in response to these changes, is constantly evolving to address new privacy threats. From the diffusion of personal computers to the explosion of the Internet, legislative efforts based on the right to be let alone concept, endeavoured to impose controls on the disclosure of data. However, the increasing number of privacy breaches provoked a shift in the European Union's perception towards a more dynamic perspective, focusing not only on the disclosure of the data but on the ongoing usage of that data, drawing elements from the right to be forgotten concept [171].

One of the pioneering attempts to legislate a data protection law, was conceived in the German state of Hessen in 1970, followed by a UK effort in 1972 and the national laws in Sweden (1973) and France (1978). These attempts focused on mitigating privacy concerns regarding the use of personal computers and the

threat of data bases withholding information on individuals [125]. In that direction, the US Department of Health, Education and Welfare circulated a report in 1973, presenting principles for fair information practices that should dictate every technological-driven use of data. The emphasis is given on reassuring individuals regarding the fair processing of their data by organisations. In essence, it is required the procurement of informed consent before collecting the data. In addition, the right of the data subjects to access their personal information in order to prohibit any unsolicited secondary use is introduced [67]. A refined version of this report was published in 1980 by the OECD, aiming to address the needs emerging from the globalisation of services that instigated discrepancies in the cross-border flow of data [95]. The aim of the report was to develop basic principles that would harmonize diverse approaches from different national privacy legislations, facilitating the data flow between countries [70].

The culmination of these efforts was the European Union's Data Protection Directive (95/46/EC) providing the principles underpinning privacy legislation in every European country [2]. The directive distinguished data in normal and sensitive, enforcing a standard level of consent pertinent to normal data and requiring an explicit consent for more sensitive data. The principles described in the directive converge with the controls identified in the literature on online privacy. More specifically, personal data should be: fairly and lawfully processed; collected for specific purposes; relevant and kept to the minimum amount necessary for the specified purpose; accurate and, where possible, kept up-to-date; and stored for a designated time when linked with the identity of the data subjects [2]. Most of the principles concentrate on the disclosure of data, remaining oblivious to the emerging needs of individuals for more dynamic types of consent.

It was not until recently, that the European Parliament authorised the General Data Protection Regulation (GDPR), a report underpinned by the same principals on which the Data Protection Directive (95/46/EC) was developed and is considered to be its replacement. Unlike its predecessor it is not a directive and will have a "direct effect" in all peers of the European Union as they are obliged to replicate it on their national legislation within two years of its agreement [80].

GDPR is a refined version as it regulates stringent procedural controls and a framework to allow individuals to exercise consent in a dynamic manner by introducing the individuals' right "to be forgotten" [80]. Regarding informed consent, GDPR bolsters its role and establishes the concept as an essential mechanism for enabling individuals to control the collection, processing and dissemination of their

personal data. It seizes the dichotomy of data to sensitive and non-sensitive data, requiring explicit and detailed consent which can be revoked at any time.

Towards revocation of consent, GDPR introduces the right to be forgotten. Drawing on and extending the existing right to keep the data up-to-date when necessary, it stipulates rights for erasing data and restrains any further use. Those rights can be exercised in five occasions: where the processing of data does not necessitate the purpose for which the data was initially collected; where consent is revoked by the data subject; when the time-duration of the legitimate given consent has elapsed; when there is an objection on behalf of the data subject; and in any encounter the procedures followed by the organisations are not compliant with the GDPR [80]. Although GDPR is pointing on the right direction, there are concerns raised regarding the applicability of revocation and how such controls can be verified and enforced.

I believe these concerns converge with the gap in the literature on revocation identified in Section 2.3.3 and highlight the necessity of an applicable model of consent and revocation. Faden and Beauchamp believe that "philosophy and law can provide a systematic approach to moral problems" [89] but when the applicability of such models is in question they cannot provide solid answers. Unlike formal methods, a discipline that can verify with a mathematical and unequivocal manner the efficiency and applicability of conceptual models in information systems.

2.5 Formal methods

Formal methods comprise of languages, tools and techniques formulated on mathematical principles that provide rigid and reliable answers regarding the specification and verification of information systems [60]. The application of formal methods on systems operating in complex and volatile environments can shed light on any inconsistencies and ambiguities that may rise in the life-cycle of the system and can provide solid ground for reasoning about its correctness [60].

The development of formal methods has revolved around two main avenues, namely formal specification and formal verification. Formal specification embraces languages structured with mathematical semantics to describe the properties of a system. A stream of literature focuses on describing sequential systems by segmenting the behaviour of the system in states which are expressed in mathematical manner as sets, relations and functions. Change in the behaviour of the system

is captured by state transitions given in the form of pre- and post-conditions. An example of such methods are Z [240] and VDM [136]. Other models draw attention to describing concurrent systems where behaviour is modelled as sequences of events and emphasis is given on whether these events could occur simultaneously. Methods such as CSP [121], CCS [180] and Temporal Logic [208] are the most dominant in the field. The value of formal specification models stem from expressing the requirements of a system in an analytical and rigorous form, enabling designers to gain useful insights, to resolve conflicts and unravel ambiguities.

Complementing formal specification, formal verification aims on corroborating a specified property of the system. Towards this direction, two different approaches have been developed, namely model checking and theorem proving. The first approach requires the design of a finite state-model, enabling designers to check whether the desired behaviour is satisfied in each or any of these states [220]. The latter approach relies on expressing both the system and the desired behaviour in a mathematical logic. The logic requires the formulation of axioms and rules that will construct the proofs. Each desired behaviour is proved when deduced by the rules and axioms of the logic [58]. Successful models combine and benefit from both approaches, such examples are [59; 156], and several tools to automate the process have been proposed and successfully implemented [110; 109; 156; 217].

Due to the often obscure mathematical models, formal methods were reluctantly embraced in information systems. Research areas that were early adopters of such methods were areas where systems' security was of paramount importance, such as military systems. In information security history, the earliest approach to achieve and verify trusted controls was access control models. The use of such models is expanded in cyberspace and the aim of the method is to determine the actions and the operations that a legitimate user can execute in a system.

An access control model consists of a subject which is the entity of the system requesting a specific access, the object which is defined as the entity of the system corresponding to the information or application that the subject may access and the action which illustrates the type of activity or operation that the subject is performing when accessing the object. Whether a subject will be allowed or not to access an object for a specific action is dictated by two components that govern the access control models, namely the access control policies and the set of control procedures. Access control policies are a set of rules that determine when an access is valid and the control procedures ensure that every access is in accordance with the policies and principles in place [225; 111]. Access control models is a fertile area

for research and as such has attracted the interest of various disciplines. Regarding formal methods, researchers have focused on developing formal models for defining and expressing access control policies, as these policies are in the heart of every access control model and a crucial factor for their success or their failure [225].

In the literature on formal methods, three different approaches have been developed to express access control policies. The first approach is the discretionary policies where every subject of the system is given authorisations and permissions (read, write, execute) for each object. These permissions are allocated to the subjects by the owners of the objects [116; 53]. Every time there is a request from a user to access an object, their credentials are matched against the specified authorizations of the object and access is only allowed if the authorisations pertained to the object specify that the user can access it. This rationale allows for greater flexibility, rendering the discretionary policies suitable for systems implemented in commercial environments. However, there is a compromise for the flexibility that they offer, as the policies focus only on the disclosure of information and are unable to handle the flow of information. In addition, there is not any control or restriction posed on data processing. For example, a subject may acquire access on data and then provide a copy to another subject to whom access to the specific data was initially prohibited [225].

The second approach was developed as a response to the limitations of the discretionary policies [225]. Mandatory policies provide a stringent and more nuanced framework on how access on the system is governed. They define a segmentation of the subjects and objects into different hierarchical security levels [34]. Instead of permitting the owners to define diverse authorisations to subjects, an approach that allows for greater flexibility and granularity, here each subject and each object is allocated to a security level that corresponds to the sensitivity of the object and the level of trustworthiness of the subject. The decisional process is underpinned by two principles that associate the security levels of subjects and objects. The first principle requires that a subject's level must supersede the level of object before access to read the information is permitted. The second principle requests that the objects' level must supersede the level of the subject before a write action is allowed. The combination of those two principles aims in prohibiting the flow of information from a high level object to a lower level, are rendering the mandatory policy approach inelastic and only applied on rigid environments, such as military systems [225]. The dominant approach of mandatory access control is the Bell-LaPadula model which defines classes forming a hierarchy from "Top secret" to "Unclassified" [34].

The limitations acknowledged in discretionary policies approach and the lack of information flow from a lower to a higher level in the mandatory policies approach fail to satisfy the requirements for services offered on cyberspace [226]. A third approach aims in providing a solution that combines the two approaches. The role-based policies approach govern access decision based on what activities users perform in the system [38]. Instead of associating each user with authorisations or defining a rigid schema of diverse security levels, the rationale in this approach is to identify different roles in the system. Each role is defined as a set of acceptable actions and authorisations related to a specific behaviour. Roles are allocated to users, who can perform one or more roles in the system, and authorisation is determined according to the role that each user adopts [224; 111]. This approach controls the usage of information as well as maintains flexibility. However, there exists a drawback since defining the appropriate roles in a system is a complex procedure [111].

2.5.1 Challenges for formal methods in privacy

A study released by the National Academies regarding "Privacy and Information Technology in a Digital Age" [258] intensifies the need to address privacy concerns and provide answers to what is technically achievable or practically impossible, in order to avoid a situation where infeasible privacy regulations are legislated. Since the applicability of different conceptual models of privacy is into question, formal methods should be embraced. The complex nature of privacy raises new challenges and new opportunities for research in the community of formal methods [251].

According to Preinbusch [214], conducting research on online privacy practices suffers from two problems. The rigid current practices opposing to consumers' behaviour that stakeholders are reluctant to abandon and the lack of models able to reason about privacy controls and their valid implementation. He concludes that the need of "advancing and integrating" formal methods is more imminent than ever, as they could shed light on understanding and supporting "consumers' privacy decision- making", provide solid mathematical proofs of proposed technical solutions and refine current practices offering "agile re-certification" based not on stakeholders claims for privacy-friendly environments but on actual implementation [214].

More specifically, formal methods can assist in the development of a framework for privacy rights, concerns and breaches. The arsenal of formal methods from state machine models to mathematical logics and model behaviour provide the appropriate tools to formalise the requirements of a dynamic and highly contextual

informational system [251]. The different conceptions of privacy can be captured by formal languages and novel logics can be applied to verify privacy-friendly system behaviour or capture the different information flow in systems required to address privacy concerns.

Privacy policies encapsulated in access control models can enable researchers to reason about privacy properties, identify ambiguities and detect inconsistencies and conflicting requirements. As highlighted in the online privacy literature review, it is important not only to control the disclosure of personal information but to control the usage and the further dissemination of the information as well. Thus, data controllers must express their policies in machine readable languages, enabling methods for reassuring data subjects that their wishes are respected throughout the life-cycle of the processing of their data. Traditional formal models cannot address these issues and as Tschantz et al. argue new logics and tools will be needed to respond to these challenges [251]. Challenges that researchers have started to investigate in the recent years.

2.5.2 Formal methods in privacy

According to Goncalves access control models must aim at safeguarding confidentiality and limited access to personal data [111]. A purpose that is in alignment with the limited access to the self privacy conception [101; 100]. The first formal models addressing privacy issues were based on traditional access control models. However there is a fundamental difference between security and privacy. In security the entity operating the system has an inherent interest in maintaining the system secure, whereas in privacy the entity who is at risk from a privacy breach is the data subject that has little control on traditional formal models. Culnan et al. [69] highlight that organisations may be operating in a secure system that can still misuse personal data raising privacy concerns. Attempts to fill this gap emphasise on embedding privacy policies and requirements mainly on role-based access control models [18; 19; 152; 182]. These models are preferred from mandatory and discretionary models because privacy policies impose restrictions not only on access but on the future use of data as well. In addition, they depend on the context and a greater flexibility is required [214; 26].

In an attempt to mitigate online privacy concerns the World Wide Web Consortium suggested the Platform for Privacy Preferences Project (P3P). P3P is an XML-based language that provides a framework for describing privacy policies [64; 63]. It

is based on the rational that the users could compare the websites' privacy guidelines with their preferences and decide whether to disclose their data or not. In particular, data controllers define P3P policies comprised of specifications regarding the required data, the purpose of collection and any conflict resolution proposals. In addition these policies impose the aforementioned restrictions in terms of data hierarchy, which is an illustration of personal attributes arranged in sets and subsets [257]. On the other hand, data subjects are provided with a policy language called "A P3P Preference Exchange Language (APPEL)" enabling them to express their privacy preferences in a compatible format [256]. A preference defined in APPEL dictates whether disclosure of data to a data controller expressing a specific P3P policy should be denied, limited or permitted. Although P3P lacks of formal specification and cannot assist in policy enforcement by ensuring that a service provider will operate as promised, it has set the threshold to develop a widely accepted language for privacy, and thus underpins most of the research in this area.

Aiming on policy enforcement, IBM developed the "Enterprise Privacy Authorization Language (EPAL)" which emphasises on formally expressing privacy policies [22; 141]. The formalisation of policies is the key factor for ensuring that policies will be implemented. EPAL requires a policy decision point responding to stimulus from a policy enforcement point. Given any action in the system, the policy enforcement point requests from the policy decision point to evaluate the appropriate EPAL policy and the answer could be either "allow, deny or don't care". In addition EPAL policies once evaluated, can create obligations, which are actions either required or prohibited in the future [22]. Similar to EPAL, another language designed to automate the process of policy enforcement is the eXtensible Access Control Markup Language (XACML) [187; 106]. Languages such as EPAL and XACML that define schemas for privacy policy languages, are the first step towards a formal methodology for addressing privacy issues. However, a "deeper integration between privacy and access control requirements is needed" [18].

Agrawal et al. [12] attempt to embed privacy policies into access control models. Their model requires expressing customers' privacy preferences in P3P format. When personal data is disclosed, the P3P expressions are captured by a policy translator and stored into a schema named "TI". These schemas contain specifications denoting under which purposes the data should be used and which entities of the system can acquire access to the data. In essence, TI schemas create restrictions that are tied with specific data and access is granted only when restrictions are satisfied [12; 158]. This approach focuses only on limiting the disclosure of personal

data [158].

Byun et al. [48] propose a role-based access control model that interprets privacy policies only as purposes for intended use of data and instead of allocating permissions to roles they assign purposes. They introduce a labelling schema, which allows the assignment of roles to either a single element, or to a single record stored in a data base, or to a column of a data base or to the entire data base table [48]. Ni et al. [191; 192] ameliorate Byun model [48] and they propose a privacy preserving role-based access control model (P-RBAC). In addition to purpose, the authors introduce a schema consisting of five privacy requirements namely data, action, purpose, condition and obligation. Thus, an action can only be implemented for the specified data if it suffices the purpose, the condition and the obligation factors. Condition is formalised as a boolean variable that captures parental and data subjects' consent and obligation is a function required to be performed either prior or after the action described in the schema [191]. A similar approach is introduced by Fischer-Hubner [93]. Instead of focusing on actions, she proposes a task-based access control model with the intention of defining the purpose of use and the indication of consent. Data can only be disclosed when a user is authorised to execute a specific task and the completion of the task depends on the disseminated data [93].

Distancing themselves from limiting the disclosure of data, some researchers adopt a different perspective that focuses on group privacy, drawing on Gavison's conception of privacy as "secrecy, anonymity and solitude"[101]. Instead of limiting access to data pertaining to a single individual, the aim is to control the type of information disclosed, by limiting data controllers' potential of deducing further information about members of a group. There is a number of techniques proposed to anonymise data [26]. In k-anonymity [244; 57] the information is suppressed until there are at least another k people in the group with similar characteristics, a rational adopted by l-diversity approach [165]. However, as Ohm argues [195], information can be either anonymised or of use. In addition, there is no guarantee that in the near future data controllers will not develop methods to unveil the hidden identity [195].

Adopting Posner's perception of privacy as the right to "conceal facts" [209], Ciriani et al. [56] introduce a language for addressing confidentiality and visibility constraints. According to their model, the data is fragmented in sets and they impose restrictions on whether the data must be visible to the user or remain confidential. There are two types of constraints expressed formally as Boolean variables. Confidentiality constraints denote that the single data set associated with

the Boolean variable or any joint sets of data should be treated as sensitive data and thus must remain invisible. Likewise, visibility constraints indicate the dissemination of a set or any joint sets of data [56].

All the aforementioned models are designed based on the limited approach to self conception or its variations and intuitively inherit its limitations as researchers focus only on restricting access to personal data. On some occasions they neglect how data is used by the data controllers while on others they neglect whether personal information is disseminated further to third parties and under which restrictions. In order to address these limitations, the focus is shifted from access control models that specify the requirements for the release of the information, to usage control models that emphasise on how data must be handled after the disclosure [273; 203]. The proposed solutions are underpinned by the conception of privacy as a set of controls and the models attempt to formalise controls identified in the online privacy literature.

One of the pivotal initiatives towards a usage control approach is Sandhu's et al. User Control Authorisation oBligation Conditions model (UCONABC) [203]. It is an approach that brings together elements from traditional access control models, trust management techniques and digital rights management (DRM) technologies. The authors claim that their approach is not "a substitute" of the aforementioned methods as they "encompass these three areas and go beyond their definition and scope" [203]. The model defines a framework that comprises of authorisations, obligations, conditions, subjects, objects and attributes. Attributes are capabilities and identities allocated to subjects and objects via an authorisation process. The authors introduce a novel characteristic of attributes as they distinguish them to mutable and immutable. Mutable attributes may change when the object is accessed or could be updated at any time by the subject whereas immutable attributes can only be changed with an administrative request. Obligations in UCONABC diverse from the definition of obligations in other access control models as they capture requirements that have to be fulfilled either prior or during the usage of data but not in the future. Conditions capture requirements emerging from the environment where the system is implemented and are not affected by subjects and objects.

The authors provide formal examples of how UCONABC can perform as a mandatory access control model, as an discretionary model and as a role-based model, while they consider privacy as a predicate dictating controls regarding access and usage. Attributes, obligations and conditions are assigned to different roles however, it is the data controller who unilaterally designs the policies that apply to

each role, reducing the expressiveness of the model. This weakness is acknowledged by the authors who, in a position paper, explore future directions of how UCON may be extended to support diverse policies for data subjects [204]. For a case study, Park et al. examine the application of UCON in social networks, a context where each user expresses different privacy preferences. Formalising users' preferences requires individual policies that would be configured by the data subject [204]. The authors fail to propose an adequate solution but highlight opportunities for future research.

For the needs of the PrimeLife project, Ardagna et al. developed a model for addressing the problem of secondary use of data [19]. The value of the model stems from focusing on policies that specify the usage and not the dissemination of data and the policy-language design comprises of data recipients, purpose of use and obligations. The proposed framework integrates the new policies focusing on defining constraints for the handling of data with traditional access control models. However they fail to provide a solution for the dissemination of those policies to other data controllers. Bussard et al. [47] address the limitation of the model and propose an XML-based language that enables data subjects to express the stream that their data will follow. In their approach the authors distinguish three different types of policies, namely "preferences" that capture access control requirements necessary for obtaining access to the data; "policies" that certify the compliant to the preferences data controllers and their intentions regarding the use of data; and the "sticky policies" that concern the dissemination of personal information between different data controllers aiming at ensuring that the preferences of the data subject will be respected by all the data controllers that may obtain their data [47].

This joint effort by Ardagna et al. and Bussard et al. set the foundations for the PrimeLife Policy Language (PPL) [20]. PPL is an extention of XACML aiming at capturing specifications regarding the usage and the further dissemination of data*. The language encompasses three different types of policies namely data handling policies that regulate the intentions of the data controller, data-handling preferences that describe the expectations of the data subject and the sticky policies that is the result of the matching process amongst the aforementioned policies. Each policy comprises of obligations and authorisations. Obligations are defined as a promise made by a data controller to a data subject in relation to the handling of his/her personal data" and are formalised in an "Event-Condition-Action manner" [206].

*The term used in the PrimeLife Project is downstream data

Authorisations are dichotomised in those describing the use of information and those defining the process of disseminating data further down the chain.

The proposed technical approach resembles that of the P3P approach. It contains a component named Policy Decision Point (PDP), which is responsible for matching data subjects' preferences with data controllers' promises and determines whether or not a party may obtain personal data, a Policy Enforcement Point (PEP), which transfers the decisions to the appropriate applications of the system and an Obligation Handler that supervises the creation and enforcement of obligations [248]. The aim of the proposed framework is not to control the flow of the data, but to prevent the undesired disclosure of information and any possible aggregation that may occur. Thus, there is no mechanism described for revoking policies.

Becker et al. propose a similar approach that emphasises on matching data subjects' wishes with data controllers' intentions, checking compliance and providing a disclosure framework for supervising the dissemination of data to other data controllers [29]. Initially developed for formalising policies that regulate the authorisation processes in decentralized systems [31; 32], SecPal is a first order logic that its signature comprises of constants and predicates describing a set of assertions. The assertions are given in the form **E says** f_0 **if** f_1**, . . . ,** f_n **where c**, where E is a constant of the logic, c is a constraint of variables and f is a fact defined as a set of predicates [30].

SecPal for Privacy (SecPAL4P) is an extension of SecPal enabling the formalisation of users' preferences and services' policies, by introducing a fix set of predicates that correspond to Personal Identifiable Information(PII) behaviour [29; 32]. Both users' preferences and service's policies are formalised by defining an upper and lower boundary. The upper bound of preferences describes how the data controllers may handle personal data, while the lower bound defines the obligations that services must comply with. In a similar manner, the upper bound of services' policies declares the possible behaviours of the data controller towards personal data, while the lower bound captures the promised behaviours. Matching a policy with a preference requires that the upper bound of a service is specified in the upper bound of the preference and the lower bound of the preference is also expressed in the lower bound of a service.

The dual expression of preferences and policies is captured formally by extending the formalisation of assertions with the formalisation of queries. Hence, the upper bound of preferences is a set of may-assertions with the form of **E says S may** f_0**, . . . ,** f_n**, where c**, while the upper bound of policies is captured by may-queries since

the data controller must request permission for handling personal data. The queries are of the form **E says S may** $f_0, \ldots, f_n$**, where c?** and are satisfied when they are a subset of the upper bound preferences. Likewise, the lower bound of preferences is captured with a will-query of the form **E says S will** $f_0, \ldots, f_n$**, where c?**, while the lower bound of policies is formalised with a will-assertion of the form **E says S will** $f_0, \ldots, f_n$**, where c?**, since the data subject requests the minimal expected behaviour prior to disclosure. The queries are satisfied when the lower bound preferences is a subset of the services' obligations [29]. Formally matching a policy to preferences, requires that both may and will queries are satisfied.

The authors further describe a framework based on checking certain assertions against the may and will queries to enforce compliance. In addition, they provide a disclosure protocol that encompasses predicates responsible for capturing users' preferences regarding the further dissemination of their data to other data-controllers [32]. Although SecPAL4P addresses changes in policies when they are still satisfied by preferences, the authors fail to describe changes in data subjects' preferences.

All the models based on the conception of privacy as controls, formalise controls mentioned in the online privacy literature regarding the collection, usage and dissemination of data. They only implicitly consider the concept of consent, however, while they provide rigid frameworks unable to capture revocation properties.

Departing from formal methods based on the conception of privacy as a form of controls, Barth et al. [26] introduce a formal language for handling the dissemination of information in situations where data controllers may operate in more than one context. Drawing on Nissenbaum's theory of contextual integrity [193], they define a temporal logic that allocates roles to people according to the context in which they operate in. Then different constraints are imposed on the disclosure of data according to the role that people perform each time. The language has successfully formalised privacy policies such as the Health Insurance Portability and Accountability Act (HIPAA) [194] and the Children's Online Privacy Protection Act [62]. However the expressiveness of the logic is limited and cannot formalise the usage of personal data.

There is a general lack of work specifically addressing the processes of consent and revocation in the context of personal data. While the concept of consent has been studied extensively in the social sciences and law, formal methods have only implicitly provided the framework of such processes. In addition, although models have advanced in expressing privacy policies regarding the collection, usage and

dissemination of data they fail to express data subjects' ability to change those policies. Individuals' privacy concerns can be mitigated when allowed to revoke or change the initially given permissions [214]. However, revocation is not related to users preferences and is usually captured in these models as invalidating a security certificate.

There appears to be a gap in the literature regarding the formalisation of a consent model as described by the law and social sciences research and the formalisation of a revocation process that will address the retrospective change of policies to data stored in information systems. A research challenge that will convert consent, revocation and data protection legislation into competitive advantage for business [214] emerges. The development of a formal privacy model able to cope with the revocation processes that stem from the dynamic privacy behaviour of online consumers can fill the identified gap.

CHAPTER 3

The conceptual model of consent and revocation

The review of the literature on privacy presented in Chapter 2 suggests that the privacy conundrum from a philosophical and legal aspect is perceived as nuanced conceptions, each of them claiming different factors to be core in determining what privacy is. The most dominant conception which understands privacy as controls, is adopted by researchers endeavouring to address privacy concerns in the online environments. Their attempt focuses on the collection, use and further dissemination of personal data while they determine different controls that once available may mitigate the risk of a privacy breach. Legislation reinforces the implementation of such controls by necessitating not only the obtaining of individual's consent for the collection of personal data but the altering and withdrawal of the initial consent given. However, the dynamic notion of the online environment renders the proposed controls rather limited for addressing privacy concerns in cyberspace while the implementation challenges raised by the recent regulatory environment [80] cannot be addressed by the existing formal models.

In most cases of online applications it is not possible for data subjects to change their consent preferences in a transparent way; without an explicit revocation capability data subjects cannot have clear and unambiguous control mechanisms to protect data privacy. When I commenced writing this thesis there was a general lack of revocation controls in social-networking, e-commerce or indeed almost any cyberspace application. Recently, social networks have introduced simple controls to revoke permissions, however the full potential of revocation controls has not been reached yet. A conceptual model for consent and revocation has still not been developed and this lack is manifested in the literature review of online privacy and

the literature review of formal methods addressing privacy problems presented in Chapter 2.

As data protection legislation has become stringent and the consequences of not conforming to it have increased, companies and governments are becoming increasingly aware that the acquisition of data subjects' consent and revocation preferences is of crucial importance. Developing a conceptual model for consent and revocation is the first step towards providing consent and revocation controls effectively. In this chapter I present a novel conceptual model of consent and revocation that considers different types of controls which an individual could give consent to, ensures that consent will be an on-going process and not an one-off event, and captures the dynamic notion of the online environment by introducing a revocation model for handling personal data in cyberspace.

The conceptual model is informed by elements highlighted in Chapter 2, since literature on privacy and data protection offers useful insights regarding the adoption and practice of informed consent in cyberspace. However, literature suggests that a series of specific issues identified in Chapter 2 emerge and references regarding revocation are so scarce that prohibits the elicitation of any concrete insights. In order to explore further the issues of informed consent and gain a holistic view of the process of revocation, the conceptual model is motivated by a series of focus-groups undertaken by the EnCoRe project [83]. The focus groups aim at understanding the control requirements for a variety of data subjects, identify possible revocation mechanisms, explore and, if possible, provide answers to what extent the data subjects are aware of the issues that emerge from the process of informed consent.

As a general observation, there is a lack of understanding of the various technical options available for implementing revocation preferences. The analysis of the focus group transcripts not only provides insights for a concrete revocation model but also gives rise to a new concept of informed revocation, which is conceived by analogy to Faden and Beauchamp's informed consent [89]. Informed revocation describes the phenomenon where data subjects tend to alter their default privacy preferences when they are informed of all the different types of revocation available to them. I argue that the novel conceptual model of consent and revocation can provide the basis for technological solutions, such as EnCoRe, that will overcome the limitations associated with informed consent. I apply the model and demonstrate its validity to a number of data-handling scenarios which have arisen in the context of the EnCoRe research project.

In Section 3.1 I argue for the necessity of the qualitative tools in order to de-

velop the conceptual model and present the methodology adopted to analyse the transcripts of the focus-groups. In Section 3.2, the novel model of consent and revocation derived by the findings of the focus-groups methodology and the existing literature on online privacy is presented. In Section 3.3, I acknowledge the different power relations between the data subjects and data controllers that affect the consent and revocation options and verify the model by applying it to diverse contexts where data subjects have different consent and revocation expectations. Section 3.4 introduces the concept of informed revocation and argue why the limitations from which the informed consent concept suffers do not apply in the process of revocation. The final Section 3.5 provides a synopsis of how this chapter contributes to knowledge.

3.1 Adopting a qualitative approach

"Qualitative research" encompasses diverse approaches that although significantly different from each other, they serve for "defining characteristics and purposes" [170]. It requires a systematic and in-depth study of people in their natural environment, avoiding assumptions made by researchers when "contriving their settings" [170] and often entailing interviews to capture experiences and perspectives of a specific concept or phenomenon. As Kaplan et al. argue "qualitative researchers assume that they do not know enough about the perspectives and situations of participants in the setting studied to be able to formulate meaningful hypotheses in advance, and instead develop and test hypotheses during the process of data collection and analysis" [140]. The generated data is in the form of transcripts and documents and must be analysed with a systematic approach, in order to retain its "inherent textural nature" which is lost "when data is quantified or aggregated" [170].

Myers [189] argues that "qualitative research differs from quantitative research most notably in the human ability to speak" and it reveals relationships without involving mathematical models [189]. Furthermore, converting "speech or contextual data" into measurable variables, will lead the participants' perception of a phenomenon and its specific "social and institutional context" to severe major losses [140].

Conducting qualitative research within the Information Systems (IS) context presupposes that the researcher understands them not as purely technical rigid systems but as soft systems shaped by their interaction with society. The aim is to

examine the interrelations of humans with the technological and organizational dimensions of the situated problem [16]. Applying qualitative research in the field of IS embraces elements from social sciences able to reason about the interplay of society with technology and offer concrete insights to researchers studying the cultural and social dimensions of a specific phenomenon [189]. Qualitative methods could provide an in depth understanding and confirm the applicability of the concepts under study.

I decided to adopt a qualitative approach to underpin, in conjunction with the literature review, the conceptual model for consent and revocation. As Maxwell declares, when a researcher studies concepts that are not easily distilled into distinct elements or is interested in the "dynamics of a process rather than its static behaviour" qualitative research is the appropriate approach to be adopted [170]. He argues that its strengths lie in the in-depth understanding of the social and cultural context in which data is generated [170]. According to Mason to study such concepts we should "interact with consumers, talk to them, listen to them and find out their options" [169]. Moreover understanding cultural influences from data generated with qualitative approach avoids complexity, as peoples' intentions and habits cannot be formed into a mathematical model.

As demonstrated in Chapter 2, individuals' perceptions of what constitutes privacy depend on the context and their cultural influences. In order to capture the cultural differences and the diverse contexts that may affect data subjects' expectations for consent and revocation controls, the EnCoRe project and more specifically Dr Edgar Whitley's group at the London School of Economics (LSE) conducted a series of focus-groups comprising participants with different backgrounds. The discussion in the focus-groups revolved around the issues of informed consent and the participants' expectations of a system that manages consent and revocation controls. I analysed the transcripts of four focus-groups to gain a deeper understanding of data subjects' wishes and a synoptic view regarding revocation mechanisms. It is worth noting that I was not present at the focus-groups and the procedure was organised by Dr. Edgar Whitley. However, the themes of the focus groups and the questions around which the discussion revolved were decided at EnCoRe general meetings, where I had the opportunity to help craft the questions and the discussed topics.

3.1.1 Focus groups as a qualitative methodology

Although the research approach that dominates the field of qualitative studies is conducting interviews to generate qualitative data, recently there is a shift to the focus group methodology as it has been acknowledged as offering a richer set of data and advantages over other qualitative approaches [268; 150; 155]. The focus-group methodology has been widely adopted for conducting research in the field of marketing in such a degree that "in practice, qualitative research has become almost synonymous with the focus group interview" [49], but it is also used in a wide range of disciplines [186]. Its application to IS and more specific in the area of privacy still remains in its infancy and usually is not a self-contained approach but is combined with other methods [186]. Therefore, analysing data generated from focus-groups, which underpin the conceptual model is a novel approach in the discipline.

The focus-groups methodology has its origins in group interviewing, an approach that has been applied for "as long as sociologists have been collecting data" [186]. The aim of focus-groups is to "focus" on a particular concept and engage all the participants in the discussion. It differs from other procedures that engage multiple participants but do not motivate interactive dialogue among them, such as Delphi groups [241]. Thus, the centre of attention is the group and not the individual [207] and whether the group will concur or conclude with recommendations is irrelevant [249].

Conducting focus groups to acquire qualitative data offers rich and detailed information regarding a specific phenomenon, concept or process. Similar to interviews, focus-groups is an interactive methodology with the advantage that during the process of collecting data there will emerge diverse viewpoints and conceptions. It is a fundamental part of the method that the researcher instead of posing questions to every interviewee, facilitates the participants to discuss between them and adopt, defend or criticise different perspectives [91; 148; 149] since "participants ask questions to each other, as well as re-evaluate and reconsider their own understandings of their specific experiences" [103]. It is this interaction and tension that offers the advantage over other methodologies and renders the approach adequate for testing conceptual models, exploring novel concepts and introducing new ideas [91]. Especially when the group dynamics are successful, participants might lead the research to "unexpected directions" [148].

The data generated from focus-groups must be the result of an organised discussion initiated by the researchers [103], which will stimulate communication amongst

the participants and it will revolve around a specific topic [148]. Acquiring data from focus-groups is particularly helpful for scrutinising participants' experiences and expectations. Regarding the conceptual model, analysing data from focus-groups can provide an in-depth understanding of peoples' expectations about consent controls and reveal preferences for revocation controls. Furthermore, the set of data can unveil not only the perceptions that people have for a particular concept but how and why they reached these perceptions. This characteristic is useful for the revocation process since via the participation in the focus-groups, "individuals can explore and clarify their views" [148] thus, providing solid ground for Kerr's psychological barriers to be further investigated and addressed in the model [147]. In addition, Morgan argues that the researcher may acquire information explaining why an issue has been salient and what characteristic results in the issue being salient [185], unveiling differences between what people say and what they do, which can be critical for examining problems regarding how well people understand the consent and revocation controls, what they wish to happen to their data and what controls they actually request.

Although group interaction can result in revealing individuals' perceptions, shifts in beliefs and indications about how people formed an opinion, there exist trades-off when adopting this approach. According to Kitzinge the group norms may prohibit voices of dissent while authoritative power relations take place as well [148]. For example, an employee may be forced to concur with his employer even if they have a different perspective. In addition the focus-groups may offer depth in the research findings but comparing to surveys and interviews the results are limited in breadth [186]. In order to overcome the downsides in the EnCoRe project, all the participants in the focus-groups were experts in their field a decision that mitigated the group norms. Regarding the depth to breadth trade-off, I sought to gain an in-depth understanding regarding the consent and revocation controls and not a wide overview of more general controls or other conceptions existing in the literature claiming to mitigate privacy concerns.

3.1.2 EnCoRe focus groups

For the EnCoRe focus groups, the decision was made to conduct a series of sessions with experts in the field of privacy and consent. Thus, the participants chosen were individuals who had already a good understanding of the complexities and challenges of the concepts to be explored. This design of focus groups is providing access into

views of professionals whose business activities shape and challenge privacy and consent policies in real life.

Finding the appropriate expert participants for each focus group was rather challenging. For the first focus group that took place in November 2008 and was held at the University of Warwick, eight participants who were representatives from civil society organisations were chosen. These participants were members of the EnCoRe's User Advisory Group. Different members of the same group participated in the second focus group that took place in January 2009. Eleven people who were data protection professionals accepted the invitation to participate. Regarding the last two focus groups, participants were selected based on recommendations and contacts known to members of the User Advisory Group and project participants. The third focus group, held also at LSE, contained nine public-sector representatives and took place on February 2009. The last focus group, was held at the University of Warwick and was part of a networking event organised by the university for a Warwickshire SME technology cluster. As a result, twenty three people participated in this focus group, whereas many of the participants knew each other well, leading to a common sharing of beliefs.

Regarding the expected outcome from the four different focus groups, the data collected from the data protection professionals was intended to provide an insight on consent practices and the various issues that organisations, in which participants work, face every day. Participants from civil society organisations sought to highlight problems that individuals are faced with and potential misuse of consent policies by organisations. The third focus group, where participants worked for the public sector, sought to investigate how one of the largest users of personal data in the UK handle data and whether consent has a key role in their practices or is complete absent (sometimes providing data to government services could be obligatory). Finally, the last focus group where employees were working for companies with limited resources, explored whether consent was considered a trivial issue or liability for these SMEs and potential problems that the implementation of a system able to provide revocation mechanisms may cause to their business.

The EnCoRe project decided to hire Tim Morley from KnowInnovation to organise and facilitate the focus group discussions. Through out the sessions Tim mentioned to the participants that he was an expert on organising focus groups rather than an expert on the concept of privacy and consent. Dr Edgar Whitley was also present as a note-taker in all focus groups, whereas Dr Nadja Kanellopoulou attended the last focus group to assist in note-taking due to the large number of

participants. Both Dr. Whitley and Dr. Kanellopoulou had minimal impact on the discussion.

The invitation sent to all participants provided information of the EnCoRe project and details of the process of focus groups. Typically, the focus groups would start with lunch where participants would meet each other and Tim. Then the session would begin with a small description of what the EnCoRe project endeavoured to achieve and a statement that there were no right or wrong answers in this process. Participants were asked to sign a consent form and were informed that the sessions were audio and video recorded (only to facilitate the transcription of the sessions) and the data generated would be available to all research staff of the project with the obligation that the transcripts created by these recordings would be anonymised. More specifically, participants were informed that:

> "the data from their session will be available to all researchers working on the project but the transcripts will be kept anonymous. The data may also be used in reports and publications and direct anonymised quotations from the transcript may be used in published output" [183].

In addition, for all focus groups a break for coffee was provided halfway through the session.

The discussion in focus groups started with an open question introduced by Tim regarding what could keep people awake at night in terms of consent issues. To ensure a flow in the conversation, participants in all focus groups were presented with various realistic scenarios in which they needed to grant and might wish subsequently to revoke consent to better control the handling of their personal data. They were asked to discuss their rational for consent and revocation achievements. More detail on the focus-groups regarding the chronological order that these took place, the number of participants attending and their background, can be found in Table 3.1.

Finally, a commercial transcription company was hired to produce the transcripts of all the focus groups. They were provided with copies of the audio and video recordings which, upon completion of the task, were destroyed. The transcripts were stored securely on the project's Twiki page, where access to all EnCoRe's members was provided. It is worth mention that the data used for the analysis in this chapter was not processed by anyone other than the author in any manner. Rather, I downloaded the transcripts from the Twiki page and used my own intuition and methods to process and analyse the transcripts.

	Date	Participants	Duration	Transcript Size
Civil society organisations	Nov 2008	8 People	3 Hours	25,500 (Words)
Data protection professionals	Jan 2009	11 People	3 Hours	27,000 (Words)
Public sector organisations	Feb 2009	9 People	3 Hours	22,000 (Words)
Regional technology cluster: SMEs	Feb 2009	23 People	1 Hour	16,000 (Words)
Total			10 Hours	90.000 (Words)

Table 3.1: Focus-groups

In the analysis section 3.3, I include relevant excerpts from transcripts in quotations. The themes that guided the discussion in the focus groups were crafted at the EnCoRe meetings. These are:

- Privacy policies
  - Are data subjects adequately informed by the privacy policies regarding the actions they are giving consent to?
  - Are privacy policies explanatory regarding the aims and procedures of the data controller?
- Controls for personal data
  - What controls the data subjects may have regarding the usage of data?
  - What do data subjects expect to happen when they refuse to give consent?
- Regulation and consent
  - How can on-going consent be achieved?
  - How can the existing regulation in UK protect the process of giving consent?

3.1.3 Content analysis as a methodology to discuss the results from focus-groups

Finding and designing an appropriate approach to generate data is not the only requirement for reaching valid conclusions in qualitative studies. Unless the data is analysed following a scientific and systematic procedure, the contextual nature of the data will be lost and the results will be likely to be unreliable. According

to Elo et al., content analysis is the appropriate methodology to adopt in order to analyse data generated from approaches that require individuals to participate in dialogues and freely express their perceptions in their own language [81]. Thus, it is one of the most widely used methodologies to analyse transcripts and documents with data generated from interviews and focus-groups.

Content analysis is a systematic research methodology applied to analyse and describe phenomena [154; 190; 126] by "designing replicable and valid inferences from texts to the context of their use" [81]. Thus, it is a scientific technique that offers insights and in-depth understanding of the concept under study. In addition, content analysis facilitates the testing of conceptual models in order to verify theories and hypotheses [81; 126]. Unlike quantitative research where the researcher compares scientific hypotheses with observed evidence, in content analysis these hypothesis are compared with inferences from the available text [81].

The origin of content analysis can be traced back to the Inquisition period in the 17^{th} century, where the diffusion of printing press forced the Church to seek different systematic approaches to identify and address "morally poor" and non-religious documents [190]. The pivotal approach of a systematic analysis occurred in Sweden in the 18^{th} century as a response to the debate that followed the publication of a ninety hymns collection, titled as the "Songs of Zion" whose author remains unknown [190]. Once the collection gained popularity, the controversy revolved around whether the hymns concealed symbols opposed to the preaching of the Church. Scholars arguing in favour of forbidding the poem from being circulated further identified possible symbols and excerpts from the collection that could potentially undermine the authority of the Church. As a response, scholars on the opposite side, conducted a quantitative analysis on other documents published by Church to highlight the existence of the same symbols in these documents. Then the debate focused on how the symbols were interpreted and in what context these emerged. This process was systematically repeated until both sides positively concurred and crystallised the interpretations of how the symbols were used in the poem and how they differed from other documents [190]. It was this response to criticism from both sides, the categorisation of concepts and the revising procedure to establish their interpretation in diverse contexts that provoked the inception of many ideas applied in the content analysis today.

Since then, the systematic analysis of texts was adopted in a wide range of disciplines and applied to establish fundamentally diverse goals. Scholars used Markov's probability theory to statistically analyse documents and propose a quantitative

approach to content analysis [190]. This approach was particularly used to examine the quality and the ethical standards of the mass increase production of newspapers in New York around 1920's. The researchers, in order to provide statistical results, were counting occurrences of specific words in the newspapers in an attempt to distinguish gossip and scandals from reporting news, while during the Second World War, content analysis was adopted to deduce information from propaganda leaflets. It was not until the end of the war that the perspective shifted from a quantitative approach to a more qualitative, as researchers embraced the method to "make use of verbally gathered data with the form of answers to open-ended interviews and focus-groups conversations" [153].

Qualitative content analysis is more refined than quantitative analysis, as it extends the counting of specific words for statistical results by focusing on the characteristics of the language and on the multiple meanings that the words may obtain depending on the context they are mentioned [260; 126]. The text is scrutinised as the researchers classify and code patterns into different categories that correspond to the multiple interpretations of the words or themes in order to "provide knowledge and understanding of the phenomenon under study" [24]. When words, themes or patterns are allocated to the same category, they are assigned the same interpretation [126]. The aim is to elicit a "broad description of the phenomenon" by developing several categories that illustrate the phenomenon from different perceptions. Elo et al. argue that this methodology is useful for developing conceptual models [81].

There exist three different approaches to facilitate the classification process, namely the inductive, deductive and summative content analysis [126]. The inductive content analysis is based on "open coding" and the categories are freely created by the researcher. In the open coding, headings and notes are written in the transcripts while reading them and different categories are created to include similar notes that capture the same aspect of the phenomenon under study [126; 81]. The process is repeated and the notes and headings are read again. The next step is to classify the categories into groups. The aim is to merge possible categories that share the same meaning [24; 126; 76]. Dey explains that this process categories data as "belonging together" and presupposes a comparison that crystallises the different perceptions [76].

The deductive content analysis requires the existence of a theory to underpin the classification process. This approach is more structured than the inductive method and the initial coding is crafted by the key features and variables of the adopted

theory [126; 81]. In the process of coding, excerpts are ascribed to categories and the findings are dictated by the theory or prior research. However, there could be novel categories that may contradict or enrich the theory by offering a refined perspective. Thus, deductive analysis is useful to validate and refine existing theories or prior research observations. According to Mayring, the existence of a theory may provide a framework to anticipate relations between categories, or predictions for the issues under study, thus helping the researcher not to deviate from the research question [172].

The summative approach is a blend of quantitative and qualitative elements. The initial step is to identify specific words of interest and examine how frequent they are used in the content. The next step embraces qualitative elements and focuses on identifying hidden meanings and different interpretations of contents [126].

The classification of categories and the interpretation of the various themes attribute a subjective element to the research and biases of the researchers may dictate the discussion and the findings of the study. In addition, content analysis is adopted in such a wide range of studies that there is not a strict process to be followed that will lead to a successful and flawless application of the methodology. On the contrary, as Elo et al. argue, it depends on the ability of the researcher [81].

Content analysis is one among a plethora of other qualitative methods to provide guidance for analysing data generated from focus-groups. Some of the other noteworthy methods include grounded theory and ethnography. Grounded theory is similar to inductive content analysis as the researcher begins from the bottom (ground) to the top, distancing herself/himself from any previous research in an attempt to arrive to a theory "grounded suited to its supposed uses" [104]. Ethnography comprises of various of methods, embracing a direct and constant communication with the agents involved. It requires the presence of the researcher for a long period of time to the environment where the phenomenon under study occurs in order to observe and gain "experience". The understanding and representation of experience shapes the findings of the research as to quote Herodotus "so far it is my eyes, my judgement and my searchings that speaks these words to you" [in [270]].

Both of these methodologies, albeit useful for their purposes, were rejected as they would exceed the needs of this thesis. Grounded theory requires the constant refinement of the questions posed to focus-groups and the presence of the researcher, while ethnography is a time-consuming methodology that may endure for more than two years in order for the researcher to obtain the appropriate experience to apply the method. The purpose of this thesis was not a social science perspective of consent

and revocation controls. Rather I decided to exploit my experience in qualitative methods and apart from the literature review, further validate the conceptual model of consent and revocation.

To analyse the data generated from the focus-groups, I applied the content analysis methodology. As there are three possible avenues to be followed regarding the classification process, a blend of inductive and deductive approaches to identify codes and develop categories was preferred. This decision was informed by the fact that there exists a well documented theory regarding informed consent and prior research to controls regarding consent is in abundance. Thus, applying a deductive approach will facilitate the validation of the conceptual consent model and provide solid grounds for its possible refinement. On the contrary, prior research regarding revocation is in infancy and to my knowledge there does not exist a theory equivalent to the theory of informed consent [10]. Thus, the inductive analysis was favoured to underpin the coding and the categories created for the conceptual model of revocation. Summative approach was rejected, as it requires the quantification of specific words and regarding revocation the word was hardly mentioned by the focus-groups since it is a word rarely used in every day language. Instead participants preferred to use synonyms such as withdrawal of consent. Thus, applying a summative approach could not have offered richer results in comparison with the other two approaches.

3.2 A novel conceptual model for consent and revocation

The blend approach provides a systematic method to analyse, code and classify the transcriptions of the focus-groups. Furthermore, it provides a crystallised view of the different interpretations that revocation may have and an in-depth understanding of how and which mechanisms may be offered to users to control the flow of their data and, as a result, mitigate their privacy concerns. In this section I present and analyse the results of the content analysis. Regarding consent, the theory is partially verified as findings suggest that it can be further extended with a new dimension. Concentrating on revocation, my initial finding was a gap between the legal and the technical perspectives. In the legal view there is an ongoing philosophical debate to understand the concept of privacy independently of technology, while computer scientists perceive privacy mechanisms only as security requirements. Even though the examined sample was relatively small, references to revocation requirements

were scant and almost without exception revocation was understood as deletion of personal data. However, the inductive approach led to a novel conceptual model of revocation and, in addition, it revealed an interesting phenomenon which I coin with the term of "informed revocation" [10].

The analysis of the data centres around five fundamental issues:

- The validity of the consent mechanisms identified in the literature that enable data subjects to control their personal data.
- The person's ability to consent to the release of her personal information and how informed the consent is.
- Mechanisms that can ensure ongoing-consent for dynamic environments.
- An individual's ability to meaningfully choose to withdraw consent.
- Defining a revocation model.

The analysis of the first three areas occurs from the deductive approach focused on the issue of informed consent while the remaining two areas are analysed with the inductive approach which investigates the issue of revocation.

3.2.1 Building the consent model

The deductive approach presupposes the existence of a theory or prior research results. The theory that underpinned the coding for the conceptual model of consent is Faden and Beauchamp's theory of informed consent [89]. In addition, I gained valuable insights from the literature on information privacy, where controls have been conceptualised mainly during the process of consent [265].

The theory of informed consent in alignment with scholars conducting research in the area of online privacy, identified controls that are applied at three different occasions:

- At the start of a disclosure.
- During the processing of data and by providing the choice for the individual to be notified.
- On personal data that is made available to others and with whom this data is shared.

Thus, I created three main categories for which a data collector requires consent from an individual:

- Collection of personal data (for storage in a database).
- Use of personal data (for analysis, processing marketing or one of many other purposes).
- Sharing or dissemination of personal data (to the public domain, or to another data collector).

Each of these cases gives rise to interesting variations and corresponding challenges. Collection can be performed in many ways (directly, indirectly), through a variety of media (an explicit registration or consent form, email, online purchases), into various forms of storage (a local enterprise server, distributed/cloud-based storage). In addition, the theory of informed consent requires the individual to acquire knowledge of how the data will be collected and of the duration of the consent given [98; 146; 89]. Another issue that emerges is that of what the Latins used to describe with the phrase "qui tacet consentit", meaning those who remain salient are assumed to give consent. Thus, consent can be considered implicit when somebody shares personal data without ever giving consent when intending to access some form of service (although in practice users are usually required to accept some terms which typically go unread, allowing a service provider to claim explicit consent when in reality it could amount to uninformed implicit consent). It is crucial for the collection process that the data subject understands the procedural practices in order to perform in an informed manner [68; 77].

Regarding the processing of data, emphasis is given on the purposes for which consent is requested [67; 4]. The purposes are practically impossible to enumerate, and no list could be exhaustive. Even if a purpose for the use of data is unambiguously defined, it is not evident how one can check that the actual use of the data matches that purpose. To illustrate the issue, consider the case of an individual giving consent for his personal data, including health records, to be used by an enterprise for medical research. The scope of "medical research" as the purpose of data collection is too broad for any realistic control to be applied to the data. The individual in question may be happy to have her data used for breast cancer research but not for diabetes research, as this may reveal private family history. In any case, there is no universal language for defining purposes clearly and unambiguously, making this aspect of consent difficult to quantify.

Providing controls for the processing of data, signifies that consent is not treated as a static procedure which is only obtained in the appropriately signified time, but it acquires an on-going element. Both in the informed consent theory and in the literature on online privacy, giving consent to a specific purpose is not sufficient for the holistic treatment of the processing of data. Further issues emerge from the secondary use of data, as aggregation practices may reveal information that individuals sought to remain private [98; 239]. Thus, it is crucial to define who may gain access to the data and what other data is available [239].

Dissemination of personal data between enterprises could cause a multitude of privacy problems, and consent for such onward sharing needs to be clearly defined and carefully enforced [242]. An enterprise may require the services of a third party to fulfil its business commitments, and in doing so share its customer database. It is up to the enterprise to ensure that the third party adheres to an adequate privacy policy, and in some cases there may be caused to be even more stringent - e.g., to prevent the third party from sharing the data onward to other parties. There are other complications also: for example, legal requirements over transferring across borders, or a situation where a public sector body needs to outsource data to a private enterprise, but is bound by tighter controls that would need to flow down through the sub-contract.

From the analysis of the transcripts, a novel category emerged. Participants in the focus-groups dichotomised the procedure of giving consent for the dissemination of data. They felt that a reasonable distinction could be made between the unrestricted sharing of data, where they felt that any sense of control is lost and the one-step further dissemination, where data could only be disseminated to a third party without that party having the ability to share the data further. In the latter case, participants felt that they were able to control the flow of the data and were more comfortable to provide consent for the dissemination of that data.

3.2.2 Building the revocation model

Unlike the deductive content approach adopted for the analysis of the conceptual model of consent, the inductive content approach is not structured and coding is freely invented by the author. Hence, classification is impacted by author's biases. The coding regarding the issue of revocation was designed based on the assumption that the revocation controls may be symmetrical to the existing controls for giving consent. The first attempt to analyse the transcripts comprised of four categories

in analogy to the four categories identified in the analysis of the consent concept. During this step, I tried to allocate excerpts of the text into the four categories. In the case were an excerpt seemed not appropriate to fit in one of the categories, I would create a novel one, based on the information from the transcript. The second step was to process again the data by focusing on recognising and merging similar categories. For example a category named as revoking data, was merged with the category deletion of data. I repeated the process several times until I crystallised the differences of various interpretations of the revocation concept. At the conclusion of the analysis the remaining categories were numbered to eight, indicating that the assumption of the apparent duality of consent and revocation does not always involve a symmetry; there exist scenarios in which consent for data to be collected or used has not been explicitly given, and yet an individual has the right to perform revocation.

Consider the case of SPAM or advertising e-mails sent to individuals' accounts without obtaining their initial consent. In such cases individuals may have a right to demand a halt, however, the mechanism for exercising that right may not be readily (if at all) available. Similarly, there are cases in which, once consent has been given, revocation may not be allowed. This is true, for instance, in the case of profiles submitted to national DNA databases in the UK (although for some, this will be conceived of as a form of forced consent or even consentless collection as they will have no option but to comply).

The analysis of data revealed that there are limited references to revocation controls and these only focused on opt-out choices. Control over personal data held regarding an individual, from the individual's point of view, can be understood as the ability to *revoke* either the data or certain permissions to process and disseminate the data, or both. Consequently, revocation has many different flavours, with subtle differences depending on how the data and the associated consent must be altered.

The principal results of the analysis is a novel taxonomy of revocation. Four fundamental types of revocation are identified(1-4 below), and another four types of revocation can be derived from the first ones (5-8 below).

1. No Revocation at all: personal data remains static, and once it has been disclosed, it is either physically impossible to revoke (how could someone ever revoke their reputation) or prohibited for various reasons (e.g., law-enforcement, data from police's DNA data-base).

2. Deletion: data is completely erased and cannot be retrieved or reconstituted

in any way. Certain privacy rights are enshrined in national and European legislation; it is worth mentioning here how our model incorporates some of the stipulations of the EU Data Protection Directive 95/46/EC [2]. In article 12, for example, the directive mentions "the rectification, erasure or blocking of data the processing of which does not comply with the provisions of this Directive, in particular because of the incomplete or inaccurate nature of the data" [2]. Rectification is a variant of revocation in the sense that a data subject may request the deletion of incorrect data held about him or herself and have it replaced with other data.

3. Revocation of permissions to process data: data subjects withdraw consent that would enable an enterprise to process or analyse their personal data for a specified purpose. EU Data Protection mentions "blocking," which corresponds exactly to revocation of permissions to process data in our model.

4. Revocation of permissions for third party dissemination: data subjects withdraw consent that would enable an enterprise to disclose information to a third party.

5. Cascading revocation is a variation on any of the above kinds of revocation, whereby the revocation is (recursively) passed on to any party to whom the data has been disclosed. Through this mechanism, data subjects are able to revoke data by only contacting the enterprise that they had disclosed their data to originally. It should be remarked that offering such a service is only practicable if data is only disclosed to organisations which themselves offer such a control.

6. Consentless revocation: personal data for whose storage and dissemination no consent has been explicitly given by the data subject, but which may need to be revoked. Again, any of the fundamental types of revocation may be invoked. This form of revocation is introduced to capture the privacy problems identified by Solove [236]. The need to revoke consentless data emerges mainly when a breach in privacy has occurred and the data subject experiences one of the acknowledged problems. For example, a picture of Jane drunk at a party was uploaded onto Facebook without her consent. As a consequence her reputation is ruined. She takes legal action in order to have the photograph removed from the site.

7. Delegated revocation: This is a kind of revocation which is exercised by a person other than the individual concerned, such as an inheritor or parent/guardian.

8. Revocation of identity (Anonymisation): data subjects may be happy for personal data to be held for certain purposes as long as it is not linkable back to them personally. Anonymisation may be regarded as a variant of revocation, in that data subjects request a change to data held so that it is no longer personally identifiable.

The last four revocation types are derivative, while the others are basic; for instance, revocation of permissions to process data may be delegated and consentless. Cascading revocation is an ideal that is difficult to implement in practice*.

3.2.3 Limitations

The conceptual model of revocation model proposed in this section may be limited in the following ways:

- The issue of granularity needs to be considered specifically for the deletion type of revocation.

- Data subjects may want to partially revoke their data, or to scramble their data instead of having it erased completely.

- The question of deletion certificates, namely, non-repudiable proofs that deletion has really been performed, but this is beyond the scope of this thesis.

- The possibility of anonymisation poses interesting problems as it makes the origin of data untraceable; there are cases where this is not in the interest of security or the common good in general. A system implementing anonymisation should have safeguards in place to ensure that data subjects will act legitimately and that the anonymised data will not impose problems on the system's integrity. On the other hand, if data is (even partly) identifiable, an enterprise can aggregate it and eventually infer to whom it refers. Such issues need to be taken into consideration when implementing revocation mechanisms.

*Indeed, one goal of the EnCoRe project is to develop an implementation of this ideal which can be deployed in actual enterprise information systems.

The novel conceptual model of consent and revocation is illustrated in Figure 3.1. The model comprises of the four fundamental types of revocation which are symmetrical to the four consent preferences depicted in the theory of informed consent and verified by the analysis of the transcripts of the focus-groups. The existence of inner circles in the figure depicts a relation between the four elements. The existence of the element illustrated in the outer circle of the Figure 3.1 presupposes the existence of all the other three elements. For example, giving consent for processing data, implies that consent for collecting data has been given already or will be given at the same time.

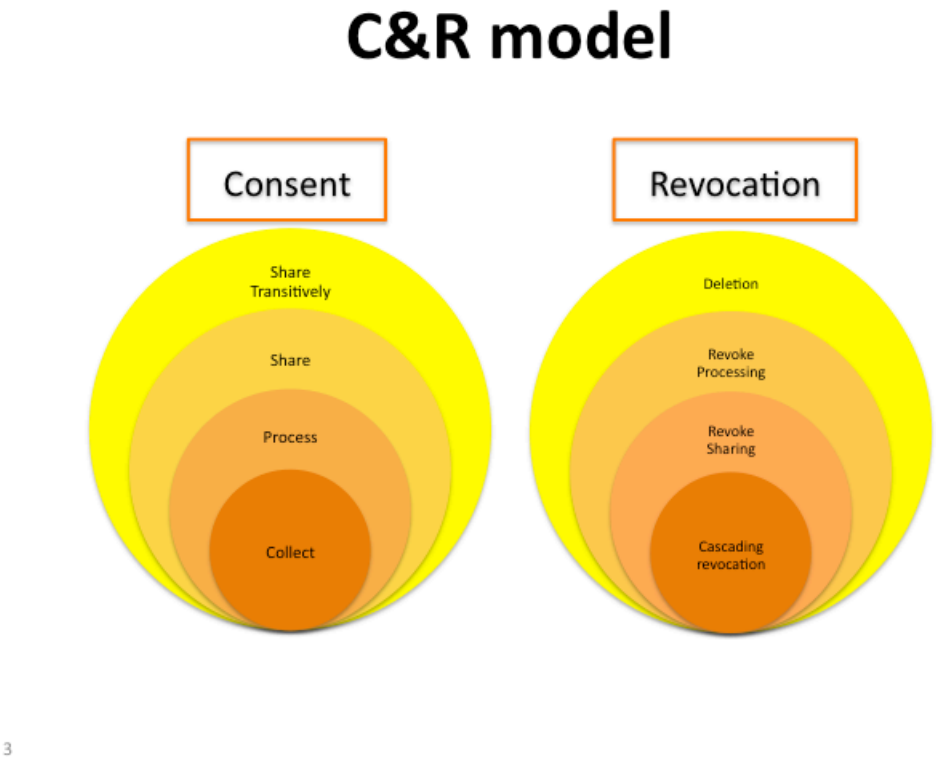


Figure 3.1: The novel conceptual model for consent and revocation

3.3 Verifying the model in different contexts

The literature suggests that privacy has a context dependent nature [174; 87; 112; 219]. The analysis of the focus-groups transcripts verified the literature findings as it emerged that the environment in which data subjects give consent to or revoke

their consent regarding the collection, use and dissemination of their personal data, drastically influences their preferences. In this section, I present the possible environments that are created, when adopting a data subject's perspective and validate further the proposed conceptual model. When stakeholders with different interests in the privacy problem interact, they establish relationships. In these relationships, there are conflicting needs to be balanced, different kinds of requirements arise and, as a result, diverse environments are formed. The concern is centred around three different categories of stakeholders:

- Data Subjects, who have a role in protecting their own personal information and specifying how it should be handled by others
- Society, which sets the standards, monitors their implementation and ensures compliance
- Data Controllers, who play a role in implementing and operating solutions

Here, I adopt a data subject's perspective, and examine the environments that are created when a data subject interacts with each one of the above three different stakeholders. Understanding the interactions that dominate in each relationship is the first step to capture the contextual nature of privacy. The focus is on interactions in order to obtain a representative view of a relationship in motion, as opposed to just a snapshot of a specific situation. Each type of interaction leads to different consent and revocation requirements and I have distinguished four cases of interest depicted in Figure 3.2 below. The arrow denotes an interaction between the data subject and a stakeholder.

Data Subject → Subject (e.g. Online Social Networks)
Data Subject → Private Sector Data Controller
Data Subject → Public Sector Data Controller
Data Subject → Society (Regulatory/Legal Environment)

Figure 3.2: The environments formed from the data subjects' interaction with the possible stakeholders of the privacy problem

The interactions of the data subject with public and private sector data controllers must be treated separately, as the participants of the focus-groups emphasised; privacy preferences of data subjects differ substantially in these two cases, as the asymmetries that emerge especially in the public related environment, create more complex situations for the data subjects to handle. Participants in the focus-groups were not asked specifically to distinguish the diverse environments in which

they perform the act of revocation. What follows below is an analysis of the identified environments. It should be noted that my decision to apply the formal model to be presented in Chapter 4, to the specific case studies that will be presented in Chapter 5, Chapter 6 and Chapter 7, was informed by the analysis of the different power asymmetries presented below.

3.3.1 Identifying data subjects' requirements

In this section I present an explanation of how privacy requirements vary across different environments.

Social-networking interactions

The social-networking environment involves interactions between data subjects mediated by a third party. The literature suggests that social networking enables data subjects to control not only their own data, but often that of their friends by providing the means to disseminate information from various data subjects to some extent [105]. Thus, data subjects are now empowered with capabilities that enable the collection process and dissemination of personal information.

In the focus-groups, there were a number of references to data subjects' interactions with other citizens, in the context of social networks. People indicated that they use sites such as Facebook and Twitter only for socialising. They do not bother to read privacy terms and conditions as they believe that the information they disclose is trivial. Even though it may be a fallacy, data subjects believe that they are always able to delete data uploaded onto these sites. They feel secure and more confident to disclose data when deletion mechanisms are in place, even though they have no guarantee that the act of deletion will actually put their data out of use. To quote one participant:

> Twitter's advanced search page allows data subjects to find deleted Tweets, an issue highlighted earlier this week after UK chat show host Jonathan Ross accidentally posted his personal email address in a message. Even though he quickly deleted the message the information was still easily obtainable, because Twitter fails to purge deleted tweets from its system.

In social networks there are some privacy controls already available. Facebook

provides fine-grained privacy settings that allow data subjects to control with whom to share what, for example [112]. Revocation in this setting is almost exclusively understood as deletion of data, and this is not always possible (as the above quote illustrates). Data subjects generally would like to have more revocation options, including anonymisation and actual deletion (expunging the data from the system altogether).

Interactions with Private Sector

When data subjects interact with private data controllers, they seek to build and enhance a relationship based on trust. Data subjects experience a lock-in effect, as they are reluctant to disclose data to another controller. They often highlight the importance of "previous experience". In contrast to the social-networks, where the interactions between data subjects have similar value for both parties (although there are exceptions [105]) because all participants share the same expectations, in this environment the situation becomes far more complex and new asymmetries emerge. These asymmetries take the form of asymmetric expectations, in which "one party expects the other party to behave in ways in which the other party does not expect or intend to behave" [105].

As mentioned above because of "expectation asymmetries," their trust is sometimes violated and data subjects wish to perform revocation mechanisms to balance the situation. Individuals are only vigilant if they happen to have experienced a breach of their privacy, and are unwilling to revoke data when the revocation mechanisms available are not clear in terms of objective and function. As stated by one participant:

> I don't really think I would actually go and pursue every company I've been shopping with and do that, because it would just be a waste, a lot of a waste of my time.

When data subjects act in this environment they mainly conceptualise revocation as deletion and opt for a regulatory organisation to certify that not only their data is properly deleted in accordance with their preferences, but also that it is not used in an arbitrary way. The importance of revocation mechanisms, understood just as deletion of data, is underlined from both data subjects and enterprises:

> I want the option [to delete my data], no matter what [damage] it does to the public [good].

I observed in the focus-groups that participants in this environment would opt for revocation mechanisms, such as revocation of the permission to process data and revocation of the permission to disseminate data. These mechanisms were not explicitly identified by them at the beginning of the focus-groups. Only through discussions at the focus-groups, where participants were incrementally forming their novel perception did they realise in how many ways they could exercise control.

Interactions with public sector

According to the literature, when data subjects interact with public data controllers new forms of asymmetries occur and thus data subject's preferences differ from the previous environments [105]. The following diverse forms of asymmetries were identified in the focus-groups:

- Asymmetry in value, in which public controllers derive high value from interactions, but data subjects derive low value
- Asymmetry in expectations, in the same sense where data subjects experience this form when interacting with a private data controller, as described above.
- Asymmetry in power, in which data subject has disproportionate ability to cause "damage" to the public controllers. Some times data subjects are forced to consent and have no information on how their data is collected, processed and disseminated among the diverse public data controllers.

From the focus-groups, participants indicated that they alter their perception of revocation when they interact with a public data controller as they experience the asymmetries presented above. The data collected and processed by the public sector is sensitive private information and citizens' interest in preventing an invasion in their private lives may be by-passed for the sake of national security, to enable medical research, or in the interest of the common good or government policy. In a focus group, a data subject expressed concern about:

> [the] merging of state and private sector, which is complicating a lot of the services under which data is actually processed, the value of data is valuable to the state for, you know, for anti-terrorist organised crime and so on and that again is making it more complicated.

Recent incidents of lost or stolen government data [253] have reduced confidence in public authorities. Data subjects are increasingly concerned about preventing arbitrary use of personal data by government services. Although data subjects acknowledge that, in particular cases, the revocation of data will not be permitted (e.g. DNA data-base), they desire revocation mechanisms so as to deal with the aforementioned problems and to restore trust in their relationship with the public sector.

Individuals are willing to share personal data for medical research if certain conditions are met. Those participated in the focus-groups have indicated that anonymity and traceability are required features of a health database if they are to disclose their medical records. However, these two concepts are in tension, often resulting in solutions based on separation but with the potential of tracing back:

> Patients - who already had the right to opt out of the scheme - now have the right to have their medical records anonymised or masked once they are put onto the system.

Due to the asymmetries, participants believed that they could not perform any revocation. However, when they realised the options that they could have, data subjects opted for the revocation of the permission to process data, to disseminate data and to delegate their decision to revoke. More specifically, when discussing scenarios with medical issues, delegated revocation was a popular option.

Interactions with society

Society could motivate enterprises to enhance their privacy mechanisms by providing revocation controls. Privacy guidelines for large enterprises exist, and law requires that these are used. Smaller enterprises need to abide by the same rules and report to the Information Commissioner Office. On the contrary, in the public sector it appears to be occasions where revocation controls are prohibited for the sake of common good [229]. In the name of society data subjects' right to privacy is invaded and in cases such as criminal records and police's DNA data bases, data may have a lifetime persistence [78].

To overview the findings in this environment, the only revocation mechanism that a data subject could apply, in terms of legislation, is the right to object to:

> unfair/unlawful processing by withdrawing the existing consent i.e. revoke and optionally replace it with a new consent; terminating any

relevant contract with the data controller/ processor; objecting on the basis that the processing is prejudicial to the data subject's "rights and freedoms" or "legitimate' interests.

Finding a balance between individuals' privacy and national security is an ongoing debate [219]. As the requirements that emerge from this environment are more of legal nature and were found also in the other three environments, this debate is beyond the scope of this model presented in the thesis, but the legal issues were assessed in the EnCoRe project.

3.4 The concept of informed revocation

There exist a lack of in-depth understanding of the different ways in which revocation can be performed and/or implemented in practice. Participants at the focus-groups perceived revocation simply as deletion of data, and they highlighted the need to be informed about the nature of deletion and the privacy protection it can actually offer. Furthermore, when they were denied the option of deletion, they were reluctant to search for alternatives. I distinguished a significant change in people's preferences when by discussing and exchanging ideas, they were informed of all the available types of revocation that they could perform in the context of a particular scenario. People become more selective and seek the revocation mechanism closest to their needs.

Tables 3.2 and 3.3 illustrate data subjects' choices of revocation mechanisms for a set of example scenarios. In Table 3.2 I have captured which revocation mechanisms data subjects expected by default. Table 3.3 shows the revocation mechanisms that data subjects chose after they were informed of their existence. It is quite evident that, once data subjects are informed of the different variants of revocation, they make more careful choices. Before being informed, they choose either to have data deleted or left intact. When given a choice between the different types of revocation identified in Section 3.2.2, they take advantage of the different controls available.

In order to explain this phenomenon, I introduce the concept of informed revocation, by analogy to Faden and Beauchamp's informed consent [89]. In their research, they argue that consent of data subjects needs to be voluntary - not the result of force or coercion - and they need to be informed about how their data is to be used, and how they can exercise rights over it if needed. When these conditions

Environments /Types	Deletion	Anonymity	Cascading Revocation	Revoke processing	Revoke sharing	Consentless revocation	No Revocation	Revoke delegation
Social networking	✓							
Medical Environment	✓						✓	
Private data controller							✓	
Public data controller	✓						✓	
Legal environment	✓						✓	

Table 3.2: Initial/Default choices

are met, consent granted for a particular use is considered informed.

I define informed revocation as a process that allows data subjects to remove and/or change permissions associated with:

- Personal data held by an enterprise.
- The purpose for which personal data may be processed by an enterprise.
- The sharing or dissemination of data by an enterprise with third parties.
- The identity of a data subject (cf. anonymisation), even for the case where consent has not been given initially.

The key characteristic of the concept of informed revocation is that the data subject should be aware of all the available types of revocation that he or she can perform, without being forced or coerced to give up any of these rights. The observed alteration in choice is also in accordance with the paradoxical behaviour that data subjects exercise when they are about to make decisions due to the psychological barriers that they experience [147]. However, if the data subjects are informed of all the possible revocation choices they may overcome the barriers presented in

Environments /Types	Deletion	Anonymity	Cascading Revocation	Revoke processing	Revoke sharing	Consent less revocation	No Revocation	Revoke delegation
Social networking	✓						✓	
Medical Environment	✓	✓		✓	✓	✓	✓	✓
Private data controller	✓		✓	✓	✓	✓	✓	✓
Public data controller				✓		✓	✓	
Legal environment						✓	✓	✓

Table 3.3: Informed Choices

Section 2.3.2 and address the issues that emerge from the decisional theory and the concept of "subjective utility" [147].

3.4.1 Differences from informed consent

The idea of consent is at the heart of codes of research ethics and the writings on that subject [89; 269]. Consent may be regarded as the opportunity to decline to take part or to withdraw from the process taking place without such decisions triggering adverse consequences for them. According to the theory of informed consent, people can only consent to something if they have received sufficient information, have understood it and have explicitly expressed agreement [89]. Its early adoption is associated with medical practice and the right of patients to be informed about the risks of medical procedures that might affect their well-being. Today its scope has broadened to include, amongst other elements, the right of online service data subjects to be informed of the way their online personal information is used.

A criticism of the concept of informed consent has been raised on the grounds that, since consent is elicited only once -before personal data is processed- it can-

not be considered "informed" throughout the lifetime of the data; in other words, consent is granted on the basis of information available at a fixed moment in time, and whether that decision may be deemed "informed" depends only on how much information was available at that moment. At a subsequent time, data might be used for alternative purposes than the data subject initially consented to, so he or she may not be fully informed.

Another concern towards achieving informed consent is how free the individuals are to participate [134]. Particularly in medical environments, people often decide to consent before they read the consent form. Patients see the process of giving consent as a mere ritual and they sign the form more as a symbolic act rather than a meaningful process that has illuminated them about the situation to be experienced.

Fisher also argues that researchers experience the same phenomenon [134]. Researchers perceive that participants share the same understanding and have the same perception about the process of consent with them and incorrectly conclude that the form they sign is informative enough for the patient to behave in an informed manner.

The revocation model in itself cannot address the criticisms levied at informed consent as a concept described in Section 2.3.2. However, I believe that an EnCoRe methodology can, and so in the project we hope to achieve informed revocation through the nature of the EnCoRe system since data subjects will necessarily engage in a process of setting consent and revocation preferences; the nature of the process tackles the problem of the non-experience of the situation. Imagine playing a game of chess where consent is equivalent to making the first move when the combination of moves are almost infinite and revocation is equivalent to deciding which move to make when the game is ending when the combination of moves could be calculated and the result predicted.

Thus, individuals are aware of the situation and do not experience the procedural misconception effect because they have already evaluated the situation and they want to exercise their right to revoke because of their experience. Furthermore, the informed revocation concept is formed in such a way that the process of revocation is unambiguous. The definitions are not open to interpretation as individuals only need to be informed of the different revocation mechanisms that they may perform and what each mechanism could achieve. However the implications that their act of revocation may have to the data controllers cannot always be predicted.

3.5 Synopsis of the findings presented in the chapter

In this chapter I illustrated the rationale for engaging qualitative methods to assist the design of a novel conceptual model for consent and revocation. The literature references on revocation are scarce and the favoured approach adopted to resolve the issue was that of focus-groups, which were organised by Dr. Whitley. I was involved in the preparation process of these focus-groups and I was allowed to acquire access to the transcripts. The methodology followed to analyse the transcripts was a blend of deductive and inductive content analysis.

Inspired by the existing theory of informed consent, I crafted the coding of consent by identifying the key elements of the theory of informed consent and the research findings in the literature. The result of this analysis was an extended model of consent as a new distinction when giving consent to disseminate data to third parties was discovered.

Regarding revocation, the literature addressing the problem of withdrawing consent is in its infancy and the inductive approach was followed. The codes were initially designed bearing in mind the dual relation between consent and revocation. However, the final model of revocation comprised of eight different guises of revocation. In addition, the inductive approach enabled the inception of the term of "informed revocation" in analogy to informed consent in order to explain the phenomenon where data subjects alter their choices of revocation when they realise the full range of controls available. The informed revocation concept does not share the same fate as informed consent regarding its limitations and is able to address the behavioural barriers that data subjects experience. It is easier for the data subjects to become informed regarding revocation because they have experience of the situation and their decisions can be more "informed" than those when they give consent.

To conclude, the novel conceptual model of consent and revocation presented in this chapter offers a holistic view of possible controls that may be offered to data subjects to control the flow of their personal information. Furthermore, the model ensures that the act of giving consent is an on-going process and not an one-off event, able to cope with the dynamic nature of the online environment. The model was validated in three different contexts, and all the requirements of the data subjects regarding privacy preferences were catered for. It is worth noting that the consent and revocation model extends the existing literature and contributes to knowledge as I am not aware of any other work that specifically addresses revocation and its

variants.

CHAPTER 4

The logic for Consent and Revocation

In Chapter 1, the paramount importance of a system providing controls to data subjects in order to express their expectations regarding the disclosure, use and further dissemination of their personal data, has been illustrated. Literature presented in Chapter 2 suggests that these controls, from an individual's point of view, can be conceived as the ability to give consent and revoke the initial consent pertaining to specific personal data. This dynamic process enables individuals to define certain privacy preferences regarding the storage, usage and further sharing of their data by the enterprises. Chapter 3 extended the concept of consent with a new distinction regarding the dissemination of data and defined different types of revocation. This resulted in a conceptual model which is the first step to address the acute need to implement practical control measures.

The second and most important step, is the design of a model capable of handling the specification process when eliciting requirements for a system offering such controls in an unequivocal manner. In that direction, embracing techniques deriving from the field of formal methods provides solid ground for the development of a mathematical model. Tools and languages in formal methods are formed on mathematical foundations and bestow rigidity and reliability when addressing the specification and verification process. Their value stems from revealing inconsistencies and ambiguities when reasoning about the correctness of a system [60].

As demonstrated in Chapter 2, the application of formal methods to privacy mainly focuses on translating privacy policies, which are mostly written in natural language, into machine readable formats [251]. Languages like P3P [64] and EPAL [22] are examples of these. Many attempts focus solely on the disclosure or usage of data, framing their solution to the boundaries of a single data controller, while other researchers emphasise either transforming personal data to make it non-

identifiable or to limit its use to a specific context [26]. In addition, a stream of literature extends models designed to capture security requirements and proposes formal frameworks for privacy preferences, focusing on the initial negotiation between individuals' preferences and business requirements without considering the life-cycle of data after a successful negotiation [32]. UCON ABC [203], in particular, is an interesting approach that combines elements from access control models and digital rights management systems to formalise authorisations, obligations and conditions pertaining to data. However, formalising privacy was not the initial purpose of the model. Thus it can not provide different policies for every user of the system, a limitation that the authors acknowledge in a position paper [204], and the concepts of consent and revocation are not explicitly formalised. An example of how the logic for consent and revocation can handle these situations is presented in Section 5.2.10.

There is a general lack of work specifically addressing the processes of consent and revocation in the context of personal data. While the concept of consent has been studied extensively in the social sciences [265], leading to work on the necessity, meaning and consequences of *informed consent* [266], few computer scientists have given the mechanics of such processes due attention. Acknowledging the gap in the literature and contributing to knowledge, this chapter presents a novel logic that will enable data controllers to more easily mitigate the risks associated with the release, handling and dissemination of personal data across information networks.

The logic is designed to allow the formalisation of the effects of expressing data subjects' consent and revocation preferences in a manner which is easily verifiable as their intention. Consent is defined to be a privacy preference when applied to personal data; the act of giving consent represents a wish for personal data to be collected, or processed, or disseminated, for a particular purpose. On the other hand, revocation is any process which corresponds to a withdrawal of consent; it is a wish for personal data to cease to be collected, processed, or disseminated. There is an auxiliary use as a testing methodology, assessing the correctness of the implementation of a system offering such preferences, can be defined based on the logic. The design of a test methodology and derivation of test suites is discussed in Chapter 8.

More specifically, the novel logic is designed in a Hoare-style manner [120] and is based on the conceptual model of consent and revocation, presented in Chapter 3. The decision to adopt Hoare triples was informed by the fact that they can express transformations of predicates. Consent and revocation is formalised as the dynamics

of access and usage and Hoare logic provides the details of what is actually happening. The logic is agnostic to the data involved, focusing on the controls that the data subject might wish to pertain to personal data.

In Section 4.1 I define the core components of the logic which comprise of actions, rights, variables and obligations, and present the formalisations of the expected actions to be performed in the system. The actions of the logic are only considered as sequential, rendering the development of a concurrency model significant for its completion. Section 4.2 explores the problems that occur when concurrent actions are triggered in the system and proposes solutions. Section 4.3 elaborates on desirable properties, called healthiness conditions, that should guide the design of the logic. Furthermore, I give evidence that the conditions hold by illustrating the state model that derives from the actions of the logic and by verifying it with a model-checker software, named Maude. There has been an incremental refinement of the logic and Section 4.5 describes this exercise which resulted in an enriched version of the logic, expressive enough to address the ambiguities encountered. The core elements of the conceptual model informed the initial attempt to formalise consent and revocation controls into a set of actions, constraints and obligations. However, when applied to the pilot Employee Data case study [11] ambiguities emerged. Section 4.5 provides an analysis of the addressed ambiguities created either by the complex notion of privacy or by the translation of natural into formal language [11] and explains the lessons learnt from the refinement process, that affected not only the development of the conclusive logic presented here, but the implementation of the EnCoRe project as well. Finally, Section 4.6 provides a synopsis of how this Chapter contributes to knowledge.

4.1 Hoare style consent and revocation logic

The logic formalises the consent and revocation processes in terms of Hoare triples. It comprises a set of rights and facts that pertain to specific data for specific principals, a set of actions appropriate to principals in the system and constraints that pertain to data expressed with variables. Rights and facts are captured with predicates that, once true, indicate that a right or a fact is true for specific data and specific actors in the system. Action is a verb, parametrised by, typically, the principal executing the action, the principal affected by it, and a package of parameters setting values for consent and revocation variables. Actions describe a transition

from one state of the system to another and denote the actors who participate in this transition. The first actor is the initiator of the action and the second is the actor influenced by this transition.The logic is parametrised by the details of the space of variables and values, although some basic ones are required to support certain actions.

The actions can be performed only when the actors involved have the appropriate rights and specific facts are true, and can either affect the rights of the principals and overwrite facts or create obligations for principals in the system, or both. Obligations capture requirements that result from applying one of the actions in the logic, denoting the necessity to apply further actions, and serve a dual purpose. They are actions that need to be triggered in the future under certain conditions (e.g., notify the data subject after five months) or actions that should be cascaded to third parties in order for the post-condition to be completed e.g., update user's data and propagate the changes to all parties that you have shared her/his data with). In the latter case, a third actor is also influenced by the transition from one state to another.

The rules of the logic are given in the form of Hoare triples, as follows:

$$\{\textit{pre-condition(rights/permissions)}\}$$
$$\textbf{action}(a, b, \delta)$$
$$\{\textit{post-condition(rights/permissions/obligations)}\}$$

The desirable initial state is captured by the pre-condition of the triple, the input that triggers the transition is defined by the action and the expected final state is described in the post-condition. Outputs from the final state, are captured with the form of obligations.

The pre-condition is a Boolean combination of (statements about) rights, facts and the values of consent and revocation variables in Φ constraints, which must be satisfied before the action can be performed. Every right consists of a sequence of three letters. The first letter denotes the actor that pertains the specific right, the second letter describes the nature of the right (right to process data or right to share data) and the third letter denotes the data that the right applies to. The constraints are expressed in variables, which constrain specific rights.

Every time an action is performed the state of the system changes from the one described in the pre-condition to the one described in the post-condition. The post-

condition, in analogy to the pre-condition, comprises of rights and permissions. In addition, it could contain obligations. In a notational contrivance, rights or predicates regarding the Φ conditions included in the pre-condition, if not negated in the post-condition, remain true. Furthermore, rights that are not mentioned in the post-condition are not affected by the action. The additional rights in the post condition will be explicitly expressed positively and when removed from it, they will be explicitly negated.

An obligation allows us to record the necessity of further actions consequent on the one just performed. The post-condition of the obligation is in addition to and should be consistent with the post-condition arising from the execution of the initial action. For example, there cannot exist a right or fact in the post-condition of the action and the negation of that right or fact in the post-condition of the obligation pertaining to the post-condition of the action, as they would be inconsistent with each other. An obligation is described as:

$$\langle \Psi_1 \rangle \ \mathbf{\Gamma} \ \langle \Psi_2 \rangle$$

and it can be interpreted as: if state Ψ_1 is satisfied, there is a requirement to apply action Γ to transit to a new state Ψ_2.

The state of the system is a set of principals, together with a set of atomic predicates relating principals with data items and, in some cases, with other principals, or with constraints giving values to consent variables. In the initial state of the system, the only rights are those of the data subject who possesses full rights over the data, while there is an assumption that both the data subject and the data controller have reached an agreement regarding the consent and revocation options*; by definition, the data subject is the 'owner' of the data and inherently possesses all the rights that are defined in the logic.

The logic is designed in such a manner that has no gap between the abstract level (that the logic describes) and the lower level (that actual programming captures). This is due to the fact that the state space as described above consists of predicates and the execution program could be the action between the pre and post conditions. The effect of executing a programming code would be to change the state presented in the pre-condition and captured with predicates, to the state expressed in the post-

*The negotiation process between the data controller and the data subject is beyond the scope of the logic and future work could embrace elements from other languages that effectively handle this process (See Chapter 9)

condition, captured in a similar vein. Thus, a program would simply assign the values of variables in each state. If, for example, $(X_1, X_2, X_3, X_4 \ldots X_n)$ are programme variables then we could envisage a formula where the $F[X_1, X_2, X_3, X_4 \ldots X_n]$ is true in a state s. The predicates described in the pre and post conditions of the Hoare triples can be mapped to the variables described in programming, leaving no gap between the abstract level and the lower level programming part.

A simple execution model is considered in the thesis, where obligations and post conditions are assumed to run to completion. As a result, the inclusion of an obligation in the post-condition of an action may be taken as an implicit shorthand for the conjunction of the post-condition of the resulting actions. It is an important healthiness condition of the logic that this should be independent of the order of the execution of the obligation actions. However, allowing obligations from arbitrary actions to interleave, may not always be possible and a more complex execution model, supporting concurrency, would be necessary. A first approach is presented in Section 4.2.1. By including the obligations given in Hoare triples in the post-condition, the logic becomes a higher order Hoare logic, since the obligations add another task to be fulfilled in order for the action of the Hoare triple to run to completion and reach the state described in the post-condition. One can imagine a function that would map a Boolean variable to a specific obligation and would return a "true value" every time the obligation is completed.

Although I believe to have captured an adequate set of actions, rights and facts to represent the consent and revocation life-cycle, I do not claim completeness when capturing the consent variables. On the contrary, new variables could be added and defined to enrich the logic's expressiveness and to capture the environment where the logic is deployed. The variables that have been used to formalise the Employee case study presented in Chapter 5 are not the same as those that have been used for the Biobank scenario illustrated in Chapter 6 or at the Identity Assurance Programme depicted in Chapter 7. In addition, the values of the variables are completely different and are customised to capture the needs of each deployment.

Legislation and business policies are considered in the logic and influence the options provided to the data subject. Since the logic is not designed to capture negotiations between entities of the system, a change in business policy or law enforcement cannot be formalised, but the result of such a change can be depicted by the logic.

I should highlight that the logic treats all data as potentially personally identifiable and therefore anonymisation is not considered. In addition, the issue of

aggregation is partially addressed. Data could be aggregated either by processing it for secondary purposes or by aggregating it with publicly available data and deriving new data. In the latter case, consent and revocation controls are not imposed on the derived data.

The syntax of the logic is described by the BackusNaur Form below. In this specification, c is a variable ranging over the actions defined in Table 4.1 P and Q are predicates ranging over the rights and facts described in Table 4.2, ϕ is a variable ranging over, but not restrained to, the variables described in Table 4.3 and x is a variable ranging over the possible data controllers captured in the system.

$$
\begin{aligned}
\langle rules \rangle &::= \{P\}c\{Q\} \\
\langle P \rangle &::= \langle right \rangle \mid \langle fact \rangle \mid \Phi \mid \pm P; \mid P_1 \wedge P_2 \mid P_1 \vee P_2 \\
\langle Q \rangle &::= \langle right \rangle \mid \langle fact \rangle \mid \langle obligation \rangle \mid \Phi \\
&\quad \mid \neg Q; \mid \pm Q; \mid Q_1 \wedge Q_2 \mid Q_1 \vee Q_2 \\
\langle obligation \rangle &::= \langle P \rangle \Gamma \langle Q \rangle \mid \forall x. \langle P \rangle \Gamma \langle Q \rangle \\
\Phi &::= \Phi \mid \neg \Phi \mid (\Phi_1 \wedge \Phi_2) \mid (\Phi_1 \vee \Phi_2) \\
&\quad \mid \Phi_1 \neq \Phi_2 \mid \Phi_1 \geq \Phi_2 \mid \Phi_1 \sqcap \Phi_2 \mid \Phi \ni \phi \\
\phi &::= \phi_t \mid \phi_v \mid \phi_\pi \mid \phi_\Pi \ldots
\end{aligned}
$$

Figure 4.1: Syntax of Hoare logic.

4.1.1 Rights, variables and actions defined in the logic

Table 4.1 presents all the actions in the logic. Table 4.2 illustrates the rights and facts of the logic and Table 4.3 depicts all the variables that are used to formalise the requirements for the three case studies.

Further to the actions listed in Table 4.1, I identify rights that will determine whether the actions can be executed. Departing from the four core elements of the conceptual model that have been captured as rights in the logic, the data subject also has the right to be notified, the right to update data and the right to delegate rights to other individuals. In addition, there exists the right of the data controller to create and store meta-data and the right of the data subject to express preferences defining the way that this meta-data will be treated. The last four rows of Table 4.2, contain predicates that deviate from the notion of the other rights. In essence, they describe facts and are used to denote that a particular situation occurs.

$\mathbf{grant}(a, b, \Phi, \delta)$	a gives consent to b to process data δ under the conditions described in the formula Φ
$\mathbf{grant}^{1}(a, b, \Phi, \delta)$	a gives consent to b to share one step further data δ under the conditions described in the formula Φ
$\mathbf{grant}^{*}(a, b, \Phi, \delta)$	a gives consent to b to share data δ transitively under the conditions described in the formula Φ
$\mathbf{grant}^{\dagger}(a, b, c, \Phi, \delta)$	a gives consent to b to share data δ with c under the conditions described in the formula Φ and this happens (c has the data)
$\mathbf{process}(b, \Phi, \delta)$	b uses data δ under Φ conditions
$\mathbf{revoke}(a, b, \Phi, \delta)$	a revokes from b permission to process data δ
$\mathbf{revoke}^{1}(a, b, \Phi, \delta)$	a revokes from b permission to share data δ one step further
$\mathbf{revoke}^{*}(a, b, \Phi, \delta)$	a revokes from b permission to share data δ transitively
$\mathbf{revoke}^{\dagger}(a, b, c, \Phi, \delta)$	a cascade-revokes permission from the third party c to whom b has shared δ with, to process data δ
$\mathbf{revoke}^{\dagger *}(a, b, c, \Phi, \delta)$	a cascade-revokes permission from the third party c to share data δ downstream, one step further
$\mathbf{delete}(a, b, \Phi, \delta)$	a requests b to delete data δ at b
$\mathbf{delete}^{\dagger}(a, b, \Phi, \delta)$	a requests b to cascade-delete data δ at b and third parties
$\mathbf{change}(a, b, \Phi', \delta)$	a changes the consent conditions for data δ to Φ' conditions
$\mathbf{update}(a, b, \delta, \delta')$	a requests to update δ with δ' and b links them
$\mathbf{update}^{\dagger}(a, b, \delta, \delta')$	a requests to update δ with δ' and b deletes the previous data δ
$\mathbf{delegate}(a, b, \Phi, \delta)$	a delegates to b under conditions Φ all rights for δ
$\mathbf{revoke\_delegate}(a, b, \delta)$	b ceases to act as a delegate on behalf of data subject a for data δ
$\mathbf{setnotify}(a, b, \Phi, \delta)$	a requires b to notify a under conditions Φ for δ
$\mathbf{setnotify}^{\dagger}(a, b, \Phi, \delta)$	a requires b to notify a under conditions Φ for δ and also to propagate these preferences to all the third parties that b has shared data δ with
$\mathbf{notify}(a, b, \delta, n^{*}, n^{\dagger})$	b notifies a that δ has been handled in accordance with n^{*} using communication method $n^{\dagger}$
$\mathbf{request}(b, c, \delta, \Phi)$	b requests from c permission to process and share data δ

Table 4.1: Meaning of the actions in the consent and revocation logic.

Right	Meaning
$aO\delta$	a owns (originates) δ
$aL\delta$	a knows (where to locate) δ
$aP\delta$	a may process δ
$aS\delta$	a may share δ (one-step further)
$aS^*\delta$	a may share δ transitively
$aN^\dagger\delta$	a may be notified for δ
$aNb\delta$	a must notify the owner of δ, b
$aU\delta$	a may update δ
$aXb\delta$	a has shared δ with b
$aR^\dagger\Phi\delta$	a must accept Φ as conditions on δ
$aR\Phi\delta$	a has accepted and will respect Φ as conditions on δ
$aXb\delta$	a has shared data δ with b

Table 4.2: List of rights and facts in the consent and revocation logic explained in English.

The variables illustrated in Table 4.3 set guards on the actions and are typically bound up in a 'preference formula' Φ. Φ can be interpreted as a stand alone predicate, expressing that the execution environment does not violate any of the constraints implied by the variables (no deadlines expired, etc). The logic is parametrised with the details of Φ.

Variable	Meaning
t	duration of consent (time-out)
u	set of purposes for which data is held
Π	parties that data may be shared with
π	parties prohibited to share data with
Π^*	parties that may acquire access to the data
π^*	parties that are prohibited from accessing the data
t^*	the number of times that data could be processed
p	binary variable that when true indicates that data is not processed by the data controller
p^*	binary variable that when true indicates that data is not processed by third parties
n^*	the means via which the notification method will be executed (email, telephone etc.)
$n^\dagger$	the notification purpose (data processed, collected, shared)
x	method to be followed when implementing deletion of data

Table 4.3: Sample of consent variables in the consent and revocation logic.

Various Φ constraints will be instantiated below, and some examples of how the

variables should be used are given. The definition of the system state, as well as the syntax of the logic, do not need to be modified to handle constraints. These are treated abstractly but a partial order $\leq$ is conceived, perhaps a constrained subset relation. Generally, there is supposed a partial order on Φ constraints, where $\Phi \geq \Phi'$ denotes that Φ' provides at least as many restrictions as Φ, so that it is no more permissive.

4.1.2 Semantics of the actions

When reading the axioms the following key should be used:

- The letter a refers to the data subject.
- The letter b refers to the data controller.
- The letter c refers to third parties, defined as data controllers who have received data from other data controllers and not from the data subject directly.

Action to capture giving consent to process data

Giving consent to process the specific data δ given by the data subject to a data controller, is formalised as follows:

$$\{aO\delta \wedge bR^{\dagger}\Phi\delta\}$$
$$\mathbf{grant}(a, b, \Phi, \delta)$$
$$\{bL\delta \wedge bP\delta \wedge bR\Phi\delta\}$$

This action is performed only by the data subject, which is formalised by requiring the $aO\delta$ right. The pre-condition is fulfilled only if the data controller is willing to respect a's conditions. The result of this action is that in the new state of the system, the rights of the data controller alter and now comprise of the right to locate the data δ and the right to process this data. Furthermore, the constraints expressed by the $bR\Phi\delta$ right ensure that the preferences of the data subject will be enforced through out data's life-cycle.

Imagine the situation where Alice (captured as a for this example) wishes to give consent to Bob (b) to process her data (δ) only for research purposes. Then the formalisation would be:

$$\{aO\delta \wedge bR^{\dagger}\Phi\delta\}$$
$$\mathbf{grant}(a, b, \Phi, \delta)$$
$$\{bL\delta \wedge bP\delta \wedge bR\Phi\delta\}$$

where Φ = purpose:u and $u \subseteq$ {research purposes}.

I introduce a different definition of the same action to describe the case where it is not the data subject but the data controller who gives consent to third parties to process data subjects' data. The reason for distinguishing the two actions is that in the later case there might be notification preferences expressed in a previous state of the system by the data subject that must be included in the post-condition. This is captured as:

$$\{bS\delta \wedge bR\Phi\delta \wedge cR^{\dagger}\Phi^{*}\delta \wedge \Phi \geq \Phi^{*}\}$$
$$\mathbf{grant}(b, c, \Phi^{*}, \delta)$$
$$\{bXc\delta \wedge cP\delta \wedge cR\Phi^{*}\delta \wedge$$
$$\langle bNa\delta\rangle\ \mathbf{setnotify}(a, c, \Phi', \delta)\ \langle true\rangle \wedge$$
$$\langle bNa\delta \wedge aN^{\dagger}\delta \wedge \Phi \ni n^{\dagger} \ni \text{“shared”}\rangle\ \mathbf{notify}(b, a, \delta, \text{“shared”}, n^{*})\ \langle true\rangle\}$$

where $\Phi' = n^{\dagger} \wedge n^{*}$.

Although the action is the same as the previous example, it is not semantically overloaded because the actors of the action are different. The only difference with the previous definition is that when this action is triggered it creates also two obligations that handle the notification process. If the data subject a has set notification controls (captured by the $bNa\delta$ right in the first bracket of the first notification) and if these conditions are met (the second bracket of the first obligation is true) then the data controller b notifies a (captured by the second obligation). The permission Φ given by a should be more permissive than the Φ^{*} given by the data controller to third parties. This rule is introduced because when the data subjects' data are disseminated to third parties, either the same or a less permissive Φ formula should be enforced to the third parties.

The action below denotes that the data subject requests from the data controller to share her/his data with a specific third party.

$$\{aO\delta \wedge bR^{\dagger}\Phi\delta\}$$
$$\mathbf{grant}^{\dagger}(a, b, c, \Phi, \delta)$$
$$\{bP\delta \wedge bS\delta \wedge bR\Phi\delta \wedge$$
$$\langle cR^{\dagger}\Phi\delta\rangle\ \mathbf{grant}(b, c, \Phi, \delta)\ \langle cL\delta \wedge cP\delta \wedge cR\Phi\delta\rangle \wedge$$
$$\langle bNa\delta\rangle\ \mathbf{setnotify}(a, c, \Phi', \delta)\ \langle true\rangle \wedge$$
$$\langle bNa\delta \wedge aN^{\dagger}\delta \wedge \Phi' \ni n^{\dagger} \ni \text{“shared”}\rangle\ \mathbf{notify}(b, a, \delta, \text{“shared”}, n^{*})\ \langle true\rangle\}$$

where $\Phi' = n^{\dagger} \wedge n^{*}$.

When this action is triggered the data controller obtains the right to process and share data and triggers an obligation to share data with a particular c. In the post-condition b possess the right to share data with any data controller he/she wishes to. In addition, there is an obligation to propagate any notification preferences to c and notify a that her/his data is shared.

Action to capture giving consent to share data only one step-further

The case where the data subject gives consent to the data controller to share his data one step further is formalised as:

$$\{aO\delta \wedge bR^{\dagger}\Phi^{*}\delta\}$$
$$\mathbf{grant^{1}}(a, b, \Phi^{*}, \delta)$$
$$\{bL\delta \wedge bP\delta \wedge bS\delta \wedge bR\Phi^{*}\delta\}$$

In the formalisation above, the data controller does not necessary possess the right to process data before she/he obtains the right to share data. The right to share data is considered to imply the right to process data, thus when an action to share data is enabled by the data subject the right to process should be obtained as well. This suggestion is a healthiness condition, explained and proven in Section 4.3. In the post-condition, the data controller possesses the right to share and the right to process the data subject to the Φ^{*} conditions.

For example, if Alice (a) is giving Bob (b) consent to share her data (δ) only with Eve and for a period of 3 months, then the formalisation would be:

$$\{aO\delta \wedge bR^{\dagger}\Phi^{*}\delta\}$$
$$\mathbf{grant^{1}}(a, b, \Phi^{*}, \delta)$$
$$\{bL\delta \wedge bP\delta \wedge bS\delta \wedge bR\Phi^{*}\delta\}$$

where Φ = parties to share data with:$\Pi \wedge$ time duration:t and $\Pi \subseteq \{\text{Eve}\}$, t= 3 months.

As in the **grant** action, I introduce a different formalisation to describe the case where the data controller gives consent to third parties to share data subjects' data. The formalisation is:

$$\{bS^{*}\delta \wedge bR\Phi\delta \wedge cR^{\dagger}\Phi^{*}\delta \wedge \Phi \geq \Phi^{*}\}$$
$$\mathbf{grant^{1}}(b, c, \Phi^{*}, \delta)$$
$$\{bXc\delta \wedge cS\delta \wedge cR\Phi^{*}\delta \wedge bR\Phi\delta \wedge$$
$$\langle bNa\delta\rangle\ \mathbf{setnotify}^{\dagger}(a, c, \Phi', \delta)\ \langle true\rangle \wedge$$
$$\langle bNa\delta \wedge aN^{\dagger}\delta \wedge \Phi' \ni n^{\dagger} \ni \text{“shared”}\rangle\ \mathbf{notify}(b, a, \delta, \text{“shared”}, n^{*})\ \langle true\rangle\}$$

where $\Phi' = n^{\dagger} \wedge n^{*}$. The rationale behind this formalisation is the same with the action **grant** for the data controller and the obligations created serve the same purpose. The difference is that in the post-condition, c has additionally obtained the right to share data.

Action to capture giving consent to share data transitively

A slightly different action is formalised below. The data subject gives consent to data controllers to share the data δ transitively†. It is captured as:

†In the logic a change of consent that needs to flow transitively to all the third parties does not happen directly by the data controller. There is an obligation created to the third parties that have shared data further down the chain

$$\{aO\delta \wedge bR^{\dagger}\Phi\delta\}$$
$$\mathbf{grant}^{*}(a, b, \Phi, \delta)$$
$$\{bL\delta \wedge bP\delta \wedge bS\delta \wedge bS^{*}\delta \wedge bR\Phi\delta\}$$

Following the same rationale, the right to share transitively implies the right to share one step further and the right to process data.The only difference with the **grant^{1}** action is that in the post condition b possesses both the $bS\delta$ and the $bS^{*}\delta$ rights.

The formalisation when the data controller triggers the action is:

$$\{bS^{*}\delta \wedge bR\Phi\delta \wedge cR^{\dagger}\Phi^{*}\delta \wedge \Phi^{*} \leq \Phi\}$$
$$\mathbf{grant}^{*}(b, c, \Phi^{*}, \delta)$$
$$\{bXc\delta \wedge cL\delta \wedge cP\delta \wedge cS\delta \wedge cS^{*}\delta \wedge cR\Phi^{*}\delta$$
$$\wedge bR\Phi\delta \wedge \langle bNa\delta\rangle\ \mathbf{setnotify}^{\dagger}(a, c, \Phi', \delta)\ \langle true\rangle \wedge$$
$$\langle bNa\delta \wedge aN^{\dagger}\delta \wedge \Phi' \ni n^{\dagger} \ni \text{“shared”}\rangle\ \mathbf{notify}(b, a, \delta, \text{“shared”}, n^{*})\ \langle true\rangle\}$$

Again the difference is that c has now the right to share data transitively.

Action to capture the processing of data

The next action captures the use of data by the data controller. Once she/he has the right to process data and the Φ conditions are respected, then the processing of data is allowed.

$$\{bP\delta \wedge bR\Phi\delta\}$$
$$\mathbf{process}(b, \Phi, \delta)$$
$$\{\neg bR\Phi\delta \wedge bR\Phi'\delta\}$$

In the post-condition, the Φ conditions according to the variables that contain may become more restrictive and the data controller must conform with the new values. Consider the case where the data subject demanded the personal data to be processed only five times. After this action, the Φ conditions should depict that

only four more times will be allowed to the data controller to use the specific data.

Action to capture changing consent choices captured in the Φ conditions

Changes in data subjects' consent choices are formalised with the action **change** by altering the initial choices captured in the Φ condition, while all other rights remain intact.

$$\{aO\delta \wedge bR\Phi\delta \wedge bR^{\dagger}\Phi^{*}\delta\}$$
$$\mathbf{change}(a, b, \Phi^{*}, \delta)$$
$$\{bR\Phi^{*}\delta \wedge \neg bR\Phi\delta \wedge \forall c.\langle bXc\delta\rangle\ \mathbf{change}(b, c, \Phi^{*}, \delta)\ \langle \neg cR\Phi\delta \wedge cR\Phi^{*}\delta\rangle\}$$

In the post condition the right $\neg bR\Phi\delta$ denotes that b no longer abides Φ for δ but is now obliged to abide Φ^{*}. Furthermore, there is an obligation for the data controller to propagate the changes to all the third parties that data have been shared with.

If for example Alice (a) requests from Bob (b) and from every party (c) that Bob has shared her data with, to change the consent and revocation permissions then the request could be formalised as follows:

$$\{aO\delta \wedge bR\Phi\delta \wedge bR^{\dagger}\Phi^{*}\delta\}$$
$$\mathbf{change}(a, b, \Phi^{*}, \delta)$$
$$\{bR\Phi^{*}\delta \wedge \neg bR\Phi\delta \wedge \forall c.\langle bXc\delta\rangle\ \mathbf{change}(b, c, \Phi^{*}, \delta)\ \langle \neg cR\Phi\delta \wedge cR\Phi^{*}\delta\rangle\}$$

where (for example) Φ = parties to share data with:$\Pi \wedge$ time duration:t
$\wedge$ purpose for processing data:u and $\Pi \subseteq$ {Eve, Sadie, Michael, Nick}, t= 5 months and u = {research purposes} and Φ^{*} = parties to share data with:$\Pi \wedge$
time duration:t $\wedge$purpose for processing data:u and $\Pi^{*} \subseteq$ {Eve, Sadie}, t= 3 months and u = {research purposes}.

In the example above the some of the values of the Φ conditions have changed to Φ^{*}. More specifically, the people with whom data could be shared with have been limited to Sadie and Eve and the time available to process this data has been reduced from 5 months to 3 months.

Action to capture delegation of consent

With the action of **delegation** data subjects may delegate their rights to other data subjects. This action can only be performed by the "owner" of the data and it is the only action that allows the $aU\delta$ right to be shared. The recipient of the right in the post-condition is the delegate. Thus, the right to update data is held only by the owner or by the delegate.

$$\{aO\delta \wedge bR^{\dagger}\Phi\delta \wedge bL\delta\}$$
$$\textbf{delegate}(a, b, \Phi, \delta)$$
$$\{bU\delta \wedge bR\Phi\delta\}$$

In this action the post-condition includes only the $bU\delta$ right. The person who holds this right is implied to also possess all the other rights available to the owner of the data. The Φ condition could describe the reason why the delegation of rights to another person took place. For simplicity reasons I will formalise the case studies introduced in the following chapters without including in the pre-conditions the $aU\delta$ right.

There is an action to formalise the revocation of delegation. With this action the delegated actor ceases to act on behalf of the data subject. The formalisation of the action is:

$$\{aO\delta \wedge bU\delta\}$$
$$\textbf{revoke\_delegate}(a, b, \delta)$$
$$\{\neg bU\delta\}$$

Actions to capture data updates

When the data subject updates [‡] the data, the initial consent rights remain the same and the only thing that changes is data. With the above description the new data is linked with the existing data.

[‡]It is assumed that all updates on data and changes on restrictions are cascaded to the third parties with whom data has been shared.

$$\{(aU\delta \vee aO\delta) \wedge bL\delta \wedge bR\Phi\delta \wedge \pm bP\delta \wedge \pm bS\delta$$
$$\wedge \pm bS^*\delta \pm bN\delta\}$$
$$\mathbf{update}(a, b, \delta, \delta')$$
$$\{bL\delta \wedge bR\Phi\delta \wedge \pm bP\delta \wedge \pm bS\delta \wedge \pm bS^*\delta$$
$$\wedge \pm bN\delta \wedge bL\delta' \wedge bR\Phi\delta' \wedge bP\delta' \wedge bS\delta' \wedge \pm bS^*\delta'$$
$$\wedge \pm bN\delta' \wedge \forall c.\langle bXc\delta\rangle\ \mathbf{update}(b, c, \delta, \delta')\ \langle true\rangle \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \textit{“updated”}\rangle\ \mathbf{notify}(b, a, \delta, \textit{“updated”}, n^*)\ \langle true\rangle\}$$

Here I make use of a notational contrivance: each predicate preceded by $\pm$ denotes that predicate or its negation, and that choice is made consistently for each predicate letter; thus if $bP\delta$ occurs in the pre-condition, then the same right should occur in the post condition both for the δ and δ' data. In the above action, the disjunction in the pre-condition allows the action to occur if the data subject($aO\delta$) or a person to whom the right $aU\delta$ has been delegated triggers the action. The principal b must have at least the $bL\delta$ because otherwise b would not possess any data to be updated. All other rights that may exist in the pre-condition, should exist in the post-condition as well.

This action does not influence rights but simply provides exactly the same rights for δ' that b has for δ. With this action, the previous data is not deleted. There is also an obligation in the post-condition to propagate the updated data to all the third parties that the data controller has shared data with. Furthermore, if there are any notification choices regarding the update of data, a notification is send to the data subject. An obligation to propagate the notification choices to third parties is not included, since the update action do not affect rights.

The data subject may request the existing data to be replaced with the new data. Notice that rights on existing data are revoked and existing data is deleted. The formalisation is:

$$\{(aU\delta \vee aO\delta) \wedge bL\delta \wedge bR\Phi\delta \wedge \pm bP\delta \wedge \pm bS\delta$$
$$\wedge \pm bS^*\delta \wedge \pm bN\delta\}$$
$$\mathbf{update}^\dagger(a, b, \delta, \delta')$$
$$\{\neg bL\delta \wedge \neg bR\Phi\delta \wedge \neg bP\delta \wedge \neg bS\delta \wedge \neg bS^*\delta \wedge$$
$$\neg bN\delta \wedge bL\delta' \wedge bR\Phi\delta' \wedge \pm bP\delta' \wedge \pm bS\delta' \wedge \pm bS^*\delta'$$
$$\wedge \pm bN\delta' \wedge \forall c.\langle bXc\delta\rangle\ \mathbf{update}^\dagger(b, c, \delta, \delta')\ \langle true\rangle \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \text{“updated”}\rangle\ \mathbf{notify}(b, a, \delta, \text{“updated”}, n^*)\ \langle true\rangle\}$$

This action does not influence any rights, rather it provides exactly the same rights that b had for data δ, with the difference that the previous data δ is deleted. Following the same rationale with the **update** action, there is one obligation in the post-condition to cascade all the updates to third parties, when the data controller has shared data with them, and another to notify the data subject regarding the update of the data.

Action to capture notification

With the action of **setnotify** the data subject is able to define the reason of the notification and the channels that the data controller should use to get in conduct with the her/him.

$$\{aO\delta \wedge aN^\dagger\delta \wedge bL\delta \wedge bR^\dagger\Phi\delta\}$$
$$\mathbf{setnotify}(a, b, \Phi, \delta)$$
$$\{bR\Phi\delta \wedge bNa\delta\}$$

where $\Phi = n^\dagger \wedge n^*$.

In this action a sets the notification preferences. Principal a can trigger the action only if he/she possess the $aN^\dagger$ right. Also in the pre-condition the principal b should at least possess the $bP\delta$ right. The more rights b possesses the more options are available for a regarding the value of the $n^\dagger$ variable (eg. only if b has the right to share data may a request to be notified when data are shared). In the post-condition, b is now aware of a's preferences (captured with the $bLn^\dagger$ and bLn^*

rights) and has the right to notify a when the conditions Φ are met.

For example, if Alice (a) would like to be notified when Bob (b) is using her data for research purposes via email, then this request can be captured as follows:

$$\{aO\delta \wedge aN^{\dagger}\delta \wedge bL\delta \wedge bR^{\dagger}\Phi\delta\}$$
$$\textbf{setnotify}(a, b, \Phi, \delta)$$
$$\{bR\Phi\delta \wedge bNa\delta\}$$

where Φ = purpose of notification:$n^{\dagger} \wedge$
means via which the notification is executed:n^{*} while $n^{*} \subseteq \{\text{email}\}$ and $n^{\dagger} \subseteq \{\text{data processed for research purposes}\}$.

The actual action when b notifies a is formalised as follows:

$$\{bNa\delta \wedge aN^{\dagger}\delta \wedge bR\Phi\delta \wedge \Phi \ni n^{\dagger} \wedge \Phi \ni n^{*}\}$$
$$\textbf{notify}(b, a, \delta, n^{\dagger}, n^{*})$$
$$\{true\}$$

If the Φ conditions, described by the $n^{\dagger}$ and n^{*} variables, are met then b notifies a (the post condition is true).

Action to capture revoking consent to process data

The options offered to data subjects that render the process of controlling personal data dynamic, are the different types of revocation. The first revoking action formalised in this section is the **revoke** action. The data subject may revoke from the data controller the right to process data.

$$\{aO\delta \wedge bP\delta \wedge \pm bS\delta \wedge \pm bS^{*}\delta \wedge bR^{\dagger}\Phi\delta\}$$
$$\textbf{revoke}(a, b, \Phi^{\dagger}, \delta)$$
$$\{bL\delta \wedge \neg bP\delta \wedge \neg bS\delta \wedge bR\Phi^{\dagger}\delta \wedge \neg bS^{*}\delta \wedge$$
$$\langle bNa\delta \wedge aN^{\dagger}\delta \wedge \Phi \ni n^{\dagger} \ni \text{“revoked”}\rangle\ \textbf{notify}(b, a, \delta, \text{“revoked”}, n^{*})\ \langle true\rangle\}$$

where $\Phi \ni p$.

In the pre-condition the only right that b should possess in order for a to trigger the action is the $bP\delta$ right. However in the post condition, the rights to share onward and transitively are revoked as well. Now b only possesses the $bL\delta$ right. If the action is triggered, then the $\Phi^\dagger$ condition should be respected (captured by the $bR\Phi^\dagger\delta$). The $\Phi^\dagger$ could be defined such as the p variable, a boolean variable capturing whether the data controller processes the data at that particular time, to be either true or false, allowing the logic to express both prospective and retrospective revocation respectively. Furthermore, there is an obligation to notify the data subject that the consent given is now revoked.

For example, if Alice (a) would like to revoke from Bob (b) the ability to process her data even if Bob is currently processing the data, then the action would be formalised as:

$$\{aO\delta \wedge bP\delta\}$$
$$\textbf{revoke}(a, b, \Phi^\dagger, \delta)$$
$$\{bL\delta \wedge \neg bP\delta \wedge \neg bS\delta \wedge bR\Phi^\dagger\delta \wedge \neg bS^*\delta \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \textit{“revoked”}\rangle \ \textbf{notify}(b, a, \delta, \textit{“revoked”}, n^*)\ \langle true\rangle\}$$

where Φ = purpose of notification:$n^\dagger \wedge$means via which notification is executed:$n^* \wedge$ data currently processed:p and $n^* \subseteq$ {email}, $n^\dagger \subseteq$ {permission to process data revoked} and p= {data processed currently}.

The data controller could request revocation of consent for processing from the third party, without the interference of the data subject. This is formalised as:

$$\{(bS\delta \vee bS^*\delta) \wedge cR^\dagger\Phi\delta \wedge bXc\delta\}$$
$$\textbf{revoke}(b, c, \Phi^\dagger, \delta)$$
$$\{cL\delta \wedge \neg cP\delta \wedge \neg cS\delta \wedge \neg bS^*\delta \wedge cR\Phi^\dagger\delta \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \textit{“revoked”}\rangle \ \textbf{notify}(c, a, \delta, \textit{“revoked”}, n^*)\ \langle true\rangle\}$$

where $\Phi^\dagger \ni p^*$.

Action to capture revoking consent to share data only one step-further

Similar to the action above, the data subject may revoke from the data controller the right to share data one step further. Following the same rationale, if the data

controller possesses the right to share data transitively, the right is also revoked.

$$\{aO\delta \wedge bP\delta \wedge (bS\delta \vee bS^*\delta) \wedge bR^\dagger\Phi\delta\}$$
$$\mathbf{revoke^1}(a, b, \Phi, \delta)$$
$$\{bL\delta \wedge bP\delta \wedge \neg bS\delta \wedge bR\Phi\delta \wedge \neg bS^*\delta \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \textit{“revoked”}\rangle\ \mathbf{notify}(b, a, \delta, \textit{“revoked”}, n^*)\ \langle true\rangle\}$$

where $\Phi \subseteq \{p, p^*\}$. Notice that the $bP\delta$ right is not revoked.

The data controller could request the revocation of the right to share data from third parties, without the action being initiated by the data subject. This is formalised as:

$$\{(bS\delta \vee bS^*\delta) \wedge cR^\dagger\Phi\delta \wedge bXc\delta\}$$
$$\mathbf{revoke^1}(b, c, \Phi^\dagger, \delta)$$
$$\{cL\delta \wedge cP\delta \wedge \neg cS\delta \wedge \neg bS^*\delta \wedge cR\Phi^\dagger\delta \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \textit{“revoked”}\rangle\ \mathbf{notify}(c, a, \delta, \textit{“revoked”}, n^*)\ \langle true\rangle\}$$

Action to capture revoking consent to share data transitively

The data subject may revoke from the data controller the right to share data transitively. This action does not affect b's rights or duties with respect to processing and sharing data one-step further.

$$\{aO\delta \wedge bP\delta \wedge bS\delta \wedge bS^*\delta \wedge bR^\dagger\Phi\delta\}$$
$$\mathbf{revoke}^*(a, b, \Phi, \delta)$$
$$\{bL\delta \wedge bP\delta \wedge bS\delta \wedge bR\Phi\delta \wedge \neg bS^*\delta \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \textit{“revoked”}\rangle\ \mathbf{notify}(b, a, \delta, \textit{“revoked”}, n^*)\ \langle true\rangle\}$$

where $\Phi \ni p$.

The data controller could request revocation of consent from the third party, without the interference of the data subject. This is formalised as:

$$\{bS\delta \vee bS^*\delta \wedge cR^\dagger\Phi\delta \wedge bXc\delta\}$$
$$\mathbf{revoke^1}(b, c, \Phi, \delta)$$
$$\{cL\delta \wedge cP\delta \wedge \neg cS\delta \wedge \neg bS^*\delta \wedge cR\Phi\delta \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \textit{“revoked”}\rangle\ \mathbf{notify}(c, a, \delta, \textit{“revoked”}, n^*)\ \langle true\rangle\}$$

where $\Phi \supseteq \{p, p^*\}$.

Actions to capture revoking consent to share and process data in a cascade way

The action where a revokes in a cascading way from c the right to process data is captured as:

$$\{aO\delta \wedge bXc\delta \wedge bR^\dagger\Phi\delta\}$$
$$\mathbf{revoke}^\dagger(a, b, c, \Phi, \delta)$$
$$\{bR\Phi\delta \wedge \langle bXc\delta \wedge cP\delta \wedge \pm cS\delta \wedge \pm cS^*\delta\rangle\ \mathbf{revoke}^\dagger(b, c, \delta)\ \langle \neg cP\delta \wedge \neg cS\delta \wedge \neg cS^*\delta\rangle \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi' \ni n^\dagger \ni \textit{“revoked”}\rangle\ \mathbf{notify}(c, a, \delta, \textit{“revoked”}, n^*)\ \langle true\rangle\}$$

where $\Phi \ni p^*$.

The subtle but important difference with the previous action is that in order for Φ to be true the data should not be currently processed by c. The rights are not revoked from the data controller. Also the cascading revocation is triggered only if the first bracket of the obligation is true. The result is that all the rights from c are revoked. The $bXc\delta$ right is not revoked in order to enable the deletion of the δ data in a later state. It is worth mentioning that this formalisation with the $bXc\delta$ in the first bracket enables the revocation of rights from b without revoking them from c since the $(bXc\delta)$ right is only revoked when c's rights are revoked. It is a function that is not possible to express with the simple logic and enriches the granularity of the different revocation actions.

The cascading revocation of the right to share data onwards or transitively is formalised as:

$$\{aO\delta \wedge bXc\delta \wedge bR^{\dagger}\Phi\delta\}$$
$$\mathbf{revoke}^{\dagger *}(a, b, c, \Phi, \delta)$$
$$\{bR\Phi\delta \wedge \langle bXc\delta \wedge cP\delta \wedge cS\delta \wedge \pm cS^{*}\delta\rangle\ \mathbf{revoke}^{\dagger *}(b, c, \Phi, \delta)\ \langle cP\delta \wedge \neg cS\delta \wedge \neg cS^{*}\delta\rangle \wedge$$
$$\langle bNa\delta \wedge aN^{\dagger}\delta \wedge \Phi \ni n^{\dagger} \ni \textit{“revoked”}\rangle\ \mathbf{notify}(c, a, \delta, \textit{“revoked”}, n^{*})\ \langle true\rangle\}$$

where $\Phi \ni p^{*}$.

Again, in the second bracket of the obligation the $bXc\delta$ right is not revoked in order to enable the revocation of the $cP\delta$ right in a later state. An action that will allow the data subject to revoke the right to share data transitively is not defined. As distinguished in the focus groups, participants felt that they had no control of their data once the data controllers shared it further. I believe that allowing third parties to share the data one step-further and not transitively does not have an impact on the data subjects belief that data is still out of their control.

Actions to request deletion of data

Deletion of data is formalised as:

$$\{aO\delta \wedge bL\delta \wedge \pm bP\delta \wedge \pm bS\delta \wedge \pm bS^{*}\delta$$
$$\wedge bR\Phi\delta \wedge bR^{\dagger}\Phi^{*}\delta\}$$
$$\mathbf{delete}(a, b, \Phi^{*}, \delta)$$
$$\{\neg bL\delta \wedge \neg bP\delta$$
$$\wedge \neg bS\delta \wedge \wedge \neg bS^{*}\delta \wedge \neg bR\Phi\delta \wedge bR\Phi^{*}\delta \wedge$$
$$\langle bNa\delta \wedge aN^{\dagger}\delta \wedge \Phi \ni n^{\dagger} \ni \textit{“deleted”}\rangle\ \mathbf{notify}(b, a, \delta, \textit{“deleted”}, n^{*})\ \langle true\rangle\}$$

where $\Phi^{*} = x \wedge p$ and Φ describes the conjunction of all Φs that the data controller should respect in the life-cycle of consent and revocation processes. At the end when the data are deleted the only right that the data controller has is the “right” to respect the conditions for deletion, thus providing a certificate that the data was deleted with the appropriate manner. Also the data controller notifies the data subject, if there is a notification choice, that the data is deleted. It is still a matter for debate how the notification could take place, because the data controller will

need the email to send the notification, thus leading to a paradox!

The only right that b should have in order for the action to occur is to know the location of data. Both in the pre-condition and in the post-condition, the $bXc\delta$ right is not revoked. It is important to highlight that if a does not revoke the data from c, then the data controller will still be obliged to propagate the "cascading deletion" action to third parties. The action is triggered only if the Φ conditions are true. Thus, the data should not be in use by b and a must specify which type of deletion should be implemented.

The data controller could request deletion of data from the third party, without the interference of the data subject. This is formalised as:

$$\{(bS\delta \vee bS^*\delta) \wedge cR^\dagger\Phi\delta \wedge bXc\delta\}$$
$$\mathbf{delete}(b, c, \Phi^\dagger, \delta)$$
$$\{\neg cL\delta \wedge \neg cP\delta \wedge \neg cS\delta \wedge \neg bS^*\delta \wedge cR\Phi^\dagger\delta \wedge \neg bXc\delta \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \textit{“revoked”}\rangle\ \mathbf{notify}(c, a, \delta, \textit{“revoked”}, n^*)\ \langle true\rangle\}$$

The action for cascading deletion is:

$$\{aO\delta \wedge bXc\delta \wedge bR^\dagger\Phi\delta\}$$
$$\mathbf{delete}^\dagger(a, b, c, \Phi, \delta)$$
$$\{\langle bXc\delta \wedge cL\delta\rangle\ \mathbf{delete}^\dagger(b, c, \delta)\ \langle \neg cL\delta \neg cP\delta \wedge \neg cS\delta \wedge \neg cS^*\delta \wedge \neg bXc\delta\rangle \wedge$$
$$\langle bNa\delta \wedge aN^\dagger\delta \wedge \Phi \ni n^\dagger \ni \textit{“deleted”}\rangle\ \mathbf{notify}(b, a, \delta, \textit{“deleted”}, n^*)\ \langle true\rangle\}$$

where $\Phi \supseteq \{x, p^*\}$.

In the pre-condition it is important to stipulate that b possesses the $bXc\delta$ right over the data δ. Thus, in the post-condition all the rights are revoked from c and also the data is deleted. It is important to highlight the subtle difference in Φ as now c should not currently process data. In addition, the $(bXc\delta)$ right is now false, indicating that the data controller has no more data shared with c.

Actions to request permission to process and share data

The action to request permission to process and share data is only used for the third case study. It is a special action due to the different business model that the project requires for the Identity Assurance Programme. The difference is that for authentication purposes the data subject gives consent to a third party to request

data from the data controller, instead of providing the data her/himself. The action is formalised as:

$$\{aO\delta \wedge bR^{\dagger}\Phi^{*}\delta \wedge cR\Phi\delta\}$$
$$\mathbf{request}(b, c, \Phi, \delta)$$
$$\{bR\Phi^{*}\delta \wedge$$
$$\langle cS^{*}\delta \wedge bR^{\dagger}\Phi\delta\rangle\ \mathbf{grant}^{*}(c, b, \Phi, \delta)\ \langle bL\delta \wedge bP\delta \wedge bS\delta \wedge bS^{*}\delta \wedge cXb\delta \wedge bR(\Phi \sqcap \Phi^{*})\delta\rangle\}$$

In the post-condition b has to comply with the restrictions posed both by the data controller c and the data subject in order to be able to process and share data. The notation $\Phi \sqcap \Phi^{*}$ define that the conditions ought to be respected by the data controller comprise of the Φ and $\Phi*$ conditions.

4.2 Concurrency issues and model

When the actions formalised in Section 4.1 are triggered simultaneously in a system offering controls captured by the logic, conflicts that need to be resolved emerge. By developing a concurrency model, I specify how actions should be handled in a concurrent manner to resolve problematic situations. It is worth noting that the semantics of the logic presented in Section 4.1, are inherently "thread-safe". Rather, the concurrency problems arise from the fact that different agents may have inconsistent views of the system caused by a number of reasons (e.g., due to the network lay).

4.2.1 Problems challenging the concurrency model

A concurrency model for the consent and revocation logic is particularly challenging for a number of reasons.

- **Rapid changing of mind:** Conceptually it must be possible for data subjects to rapidly change their mind regarding their consent and revocation preferences, which means that the concurrency model must impart some ordering on actions to ensure that the notion of the most recent set of preferences can be maintained correctly.

- **Identity of data:** The data subject and data controllers may have subtly different understandings of what constitutes an item of data to which controls must be applied.

- **Different preferences for different data controllers:** It must also be possible for data subjects to choose differing consent and revocation preferences (for the same data) for each data controller that they interact with, after all that is their prerogative.
- **The unique case of deletion:** The logic should provide sufficient evidence that situations where the conflicting preferences include a deletion request are handled.
- **Circular chains of actions:** If someone has to consider an ecosystem of data controllers, then she/he has to allow for the creation of supply chains of data sharing and associated service provision which can have circular connectivity.

4.2.2 The rapid changing of mind

The problem of rapid changing of mind occurs when the data subject decides to change their consent choices so quickly that the system is not able to implement the first change before the action for the next change occurs. Thus, there is a risk that the former action might not be implemented (or it could be implemented later than the second action) and the final result could be different from what the data subject expects.

For example, suppose the data subject decides data to be shared with third parties but regrets it soon after and decides to revoke that action by deleting data from the third parties. If the decision to delete data is so rapid that the second action in the system occurs before the sharing of data is implemented by the system, then in the final state of the system, the third parties will possess data from the data subject (in essence the sequence of the actions is reversed).

To avoid such issues a concurrency model that will highlight which sequence of actions could create problematic situations must be defined and a resolution policy should be provided. What safely can be argued is that:

- All grant actions could run in parallel. The pre-conditions of the actions are disjoint and the post conditions are compatible (i.e. do not interact).
- All revocation actions could run in parallel for the same reason that the grant actions run in parallel.
- A single notify, a single change and a single update action could run in parallel,

as they affect different rights and Φ restrictions. However, two of the same kind cannot.

Since the consent and revocation logic is designed based on a user-centric perspective, the actions that derive from the data subject will always have the highest priority. The obligations that occur from the data subject's wishes have the second highest priority, whereas the actions that are triggered by the data controller have the lowest priority. No actions can be triggered by third parties that could influence the state of the data controller's system.

Based on the above rationale and looking the system from a global view (i.e. someone who can see everything), the concurrency model is as follows:

- When the grant, notify, change and update actions are triggered in parallel (denoted by the $\|$ symbol) then the grant action is executed first and then the other actions are all executed simultaneously.
  - grant(a,b,δ)
  - notify $\|$ change $\|$ update
  - If the grant action is triggered by the data controller and the other actions are triggered by the data subject, then the priority is reversed.
    - notify $\|$ update $\|$ change
    - grant(b,c,δ)
- When the revoke, notify, change and update actions are triggered in parallel then the revoke action is executed first and then the other actions are all executed simultaneously.
  - revoke(a,b,δ)
  - notify $\|$ change $\|$ update
  - If the revoke action is an obligation, triggered by the data controller and the other actions are triggered by the data subject, then the priority is reversed.
    - notify $\|$ update $\|$ change
    - revoke(b,c,δ)
- When the actions revoke, grant update, change and notify are triggered together then:

  - The actions with the highest priority are the "*grant*" and "*revoke*" actions. The actions then are ordered by using a time stamp and are processed sequentially. The action with the least recent time stamp will be triggered first, followed upon completion from the second action.
  - The second action is triggered (either grant or revoke).
  - All the other actions are triggered simultaneously.
  - If a revoke action is an obligation, then the obligation is triggered first, the update, notify and change actions next and the last action to be triggered is the grant action.
  - Based on the same rationale, if the grant action is triggered by the data controller, the revoke action has the highest priority. The grant action will be triggered first, then the revoke action is triggered and the last actions to be executed are the update, notify and change actions.
  - If the grant action is triggered by the data controller and the revoke action is an obligation, then the update, change and notify actions have the highest priority. Thus, the first action to be triggered is the grant action by the data controller, the second should be the obligation to revoke and the last actions are the update, change and notify actions.

- In the case where grant, revoke, change, update and notify actions are triggered by the data subject, revocation, change, update and notify actions are triggered by obligations and grant and revoke actions are triggered by the data controller, then the priority is:

  1. Resolve the time issue between the grant and the revoke actions that were triggered by the data controller and trigger the first action.
  2. Resolve the time issue between the grant and the revoke actions that were triggered by the data controller and trigger the second action.
  3. Resolve the time issue between the revocation or grant action that is triggered by an obligation and trigger the first action.
  4. Resolve the time issue between the revocation or grant action that is triggered by an obligation and trigger the second action.
  5. The actions update $\|$ notify $\|$ change triggered by an obligation.
  6. Resolve the time issue between the grant and revoke actions triggered by the data subject and trigger the first action.

7. Resolve the time issue between the grant and revoke actions triggered by the data subject and trigger the second action.
8. The actions update ‖ notify ‖ change triggered by the data subject.

4.2.3 The special case of deletion

The case of deletion is one aspect of the problem presented in Section 4.2.2. If the data is deleted then no more actions could have an effect in the system. The solution proposed for the rapid change of mind effectively addresses the issue of deletion.

4.2.4 Identity of data

The logic handles the consent and revocation preferences in a system. I assume that the data subject's and the data controller's perceptions are in alignment, regarding the policies of the system. The right $bR^{\dagger}\Phi\delta$ serves this purpose. Thus, the assumption is that data subject and the data controller share a common understanding of what data item is. Any negotiations between the data subject and the data controller could be formally described with logics complementary to the logic, such as SecPal [32] and create opportunities for future research which are presented in Chapter 9.

4.2.5 The circular problem

When the unrestricted propagation of data is allowed, it could potentially lead to a situation were a data controller will propagate obligations to the third party and the third party will propagate the same obligations to the data controller, thus creating an infinite loop of obligations.

To avoid this situation, a possible solution will demand the restriction of the propagation of data. A third party should only accept data from a data controller or another third party, if she/he does not already possess the same data from the same data controller. The collection of the same data is allowed so many times as the number of the existing data controllers registered in the system. Data controllers are considered those parties that have acquired data directly from the data subject. This rule will enable the logic to describe situations where the data subject has expressed different consent and revocation preferences for the same data to different data controllers. In addition a hash list could be created when the data is stored to record

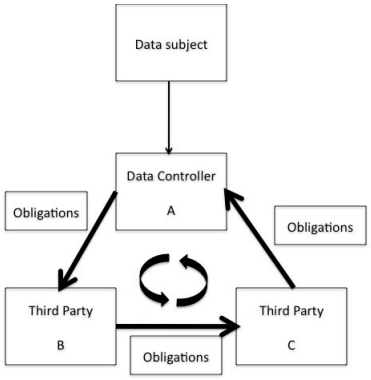


Figure 4.2: The circular problem

restrictions posed by the senders of data. This hash list allows different consent and revocation preferences for the same data without propagating obligations.

In Figure 4.3 the sharing of data from C to A is prohibited in order to tackle the infinite propagation of obligations. There is only one data controller, so the data could be stored only one time.

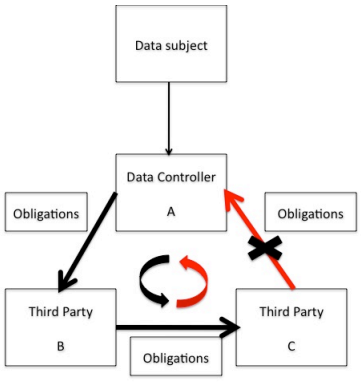


Figure 4.3: Actions prohibited by the proposed solution

In Figure 4.4 the sharing of data from C to A is allowed since there exist two different data controllers and the entity A has one copy of data. The data controller will determine which data to use, based on whose behalf the data is processed.

4.2.6 Different preferences for different data handlers

The ability to set different preferences for different data handlers regarding the same data is one aspect of the problems described in Section 4.2.5. It should be

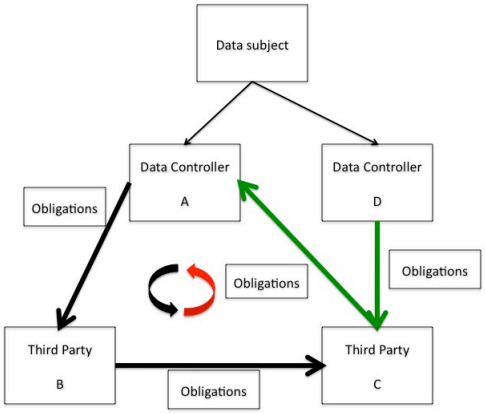


Figure 4.4: Actions allowed by the proposed solution

highlighted that only the data subject may choose completely different preferences for the same data to different data controllers. Once these choices are set, then the data controllers could only share the data further by posing the same or more restrictive controls on the data. Once the data is received there is a list created to keep track of the different constraints that should be respected by different parties. Figure 4.5 below illustrates how the data handler D processes data for purpose 1 when the procedures concern data handler B and processes data for purpose 2 when the procedures concern the data handler C.

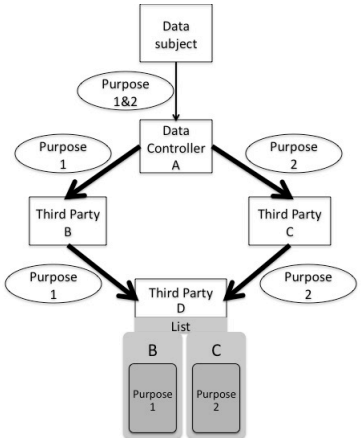


Figure 4.5: List that contains different constraints for different parties

4.3 Healthiness conditions

In the logic the actions are defined in a non-deterministic behaviour to describe transitions in the system and the conjunction of predicates in every pre and post condition indicates different states of the system. Furthermore, these predicates could be constrained by variables comprising the Φ conditions. It is essential to ensure that what is written in logical expressions is achievable, meaning that there do not exist described behaviours that cannot be allowed by the system. In addition, there are desirable behaviours and properties that are expected to be captured by the logic. In order to address these issues, I define several healthiness conditions as desirable emergent properties of the logic.

A systematic approach to define the healthiness condition, is followed. Departing from the initial state of the EnCoRe system for a specific data controller I examine what actions are feasible from that state, what predicates describe the reachable states mediated by these actions and which variables could constrain these predicates. The same rationale is applied to all the states of the system and the healthiness conditions are defined in order to connect the logical transitions with the desired behaviour of the system in the real world. Counter-examples are used to identify undesirable behaviour.

Healthiness conditions, albeit desirable, are irrelevant if they are not satisfied by the models. In order to prove the healthiness conditions of the logic, I created a state-model that comprises of all the different states that the system might reach for a single data controller and for a single data. All the actions formalised in the logic and the transitions they describe are included in the state model. The states are defined according to the pre and post conditions of the actions. Figure 4.6 illustrates all the "grant and process" actions that are defined in the logic and the transitions that these actions depict, while, in a similar manner, Figure 4.7 presents all the "revoke and delete" that can take place in the system. Figure 4.8 describes the actions that are independent, thus their transitions does not interfere with the transitions of the "grant and revoke actions".

The proofs are performed using Maude [247]. Maude, is a "high-performance reflective language and system supporting both equational and rewriting logic specification and programming" [247] applicable to a variety of contexts. Rewriting logic in Maude was preferred because, in concurrent change, the software can "naturally deal with state and with concurrent computations" [247]. The healthiness conditions that the logic satisfies are presented in the next section, while the proofs developed

in Maude can be found in Section 4.4. All the healthiness conditions relating to the characteristics of the actors are proven in Section 4.4, while only the fifth condition relating to the characteristics of the system is proven in the same section. The remaining conditions either are axioms of the logic or are desirable properties. Table 4.4 explains how the different states in the Hoare logic are captured in Maude.

	Actions	States in Maude	Rights in the logic
1	grant(a,b,Φ, δ)	O	aOδ
2	grant^1(a,b,Φ, δ)	P	bPδ
3	grant*(a,b,Φ, δ)	P,S	$bP\delta \wedge bS\delta$
4	any action providing data to third parties	P,S,S*	$bP\delta \wedge bS\delta \wedge bS^*\delta$
5	delete(a,b,Φ, δ)	X,P	$bXc\delta \wedge bP\delta$
6	revoke(a,b,Φ, δ)	X,P,S	$bXc\delta \wedge bP\delta \wedge bS\delta$
7	delete†(a,b,Φ, δ)	X,P,S,S*	$bXc\delta \wedge bP\delta \wedge bS\delta \wedge bS^*\delta$
8	revoke*(a,b,Φ, δ)	X	bXcδ
9	revoke^1(a,b,Φ, δ)	Current State (CS)	Any rights of the simple model
10	any action revoking rights from third parties	$CS + Notifications$	$aN^\dagger\delta \wedge aNb\delta$
11	process(b,Φ, δ)	$CS + Updates$	$aU\delta$
12	change(a,b,Φ', δ)	CS + Notifications + Updates	$aN^\dagger\delta \wedge aNb\delta \wedge aU\delta$
13	setnotify(a,b,Φ, δ)		
14	notify(a,b,, $\delta, n^*, n^\dagger$)		

Table 4.4: Explanation of the state-machine's notations

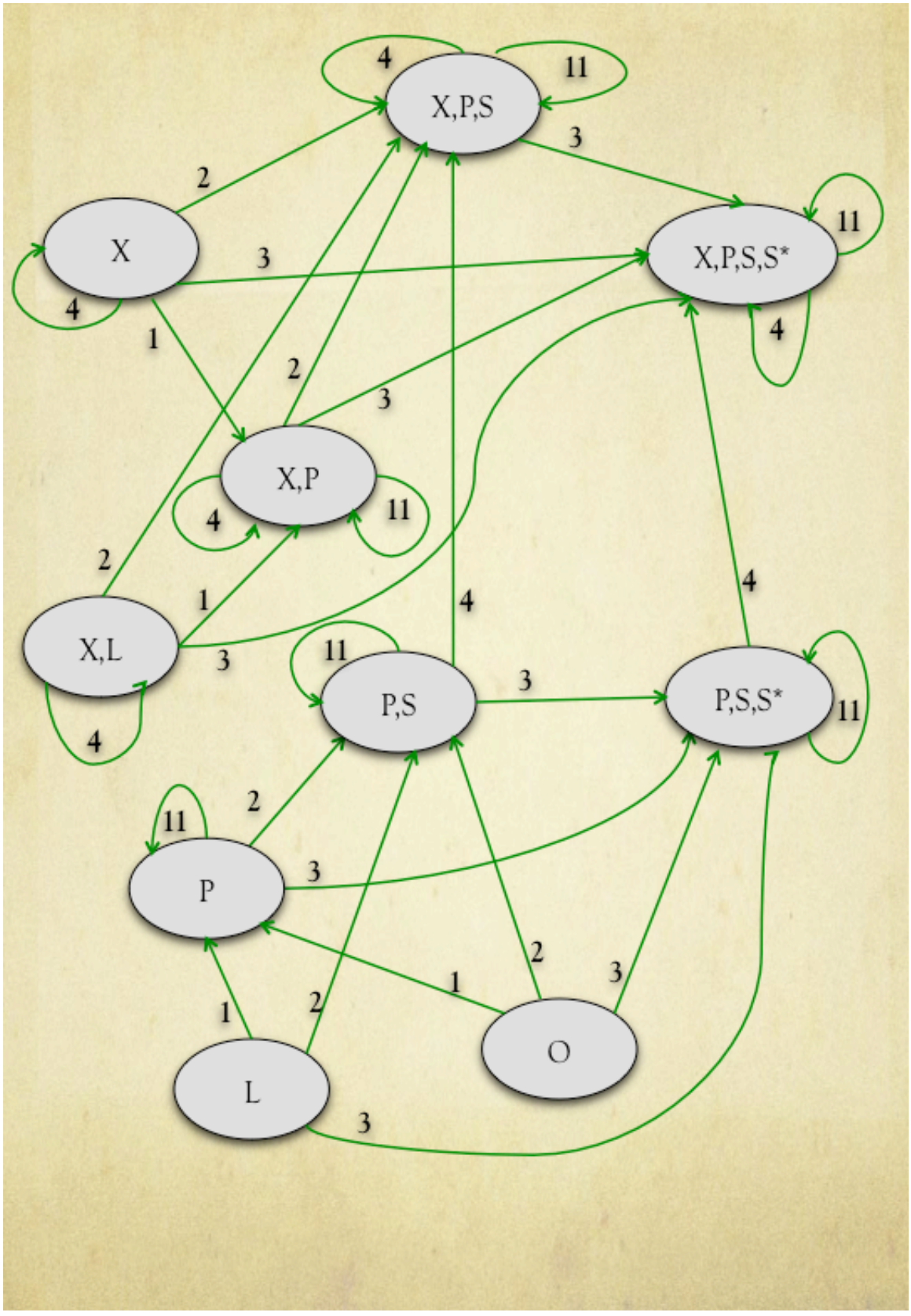


Figure 4.6: The state machine model depicting only the "grant and process" actions examined in Maude

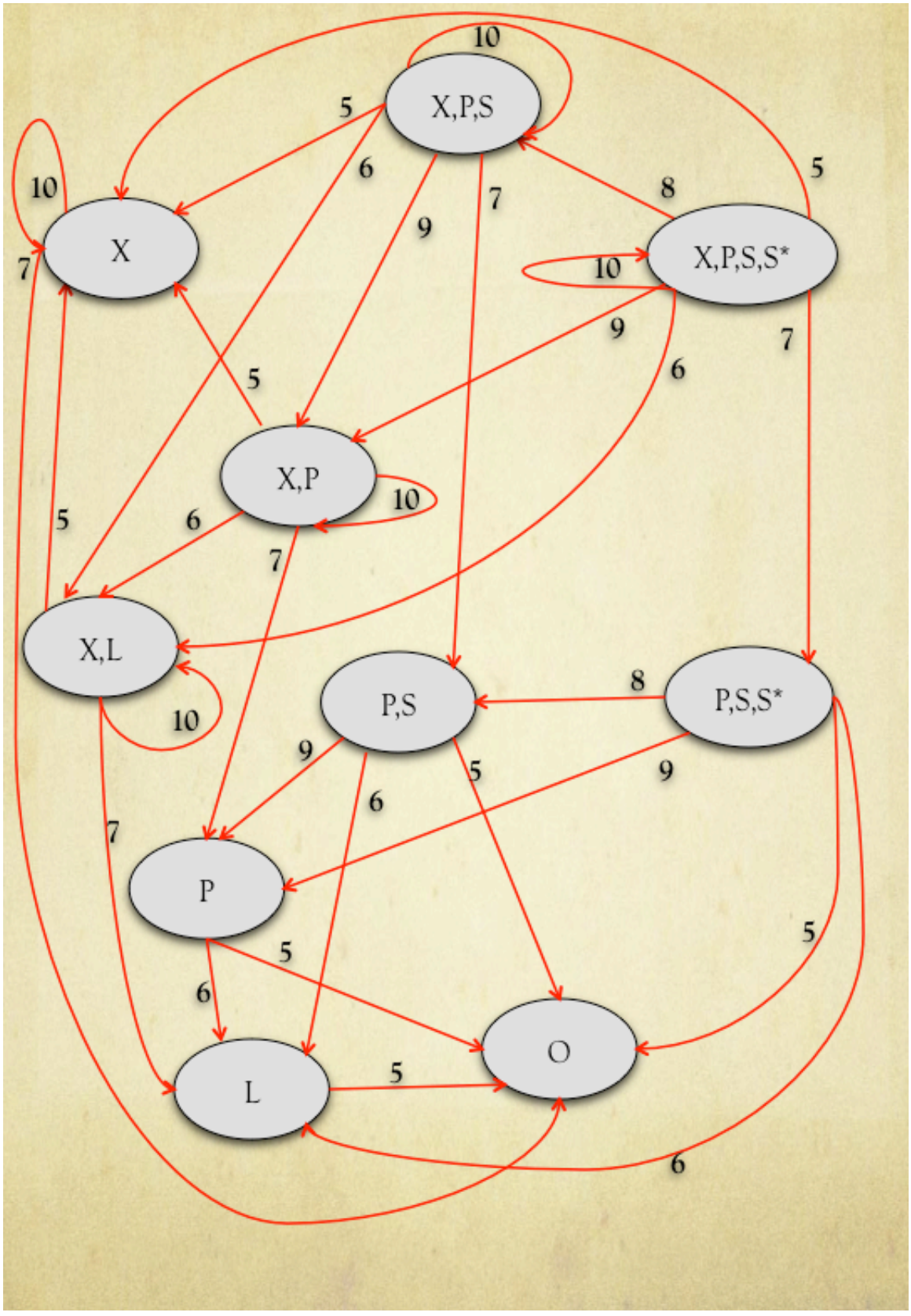


Figure 4.7: The state machine model depicting only the "revoke and delete" actions examined in Maude

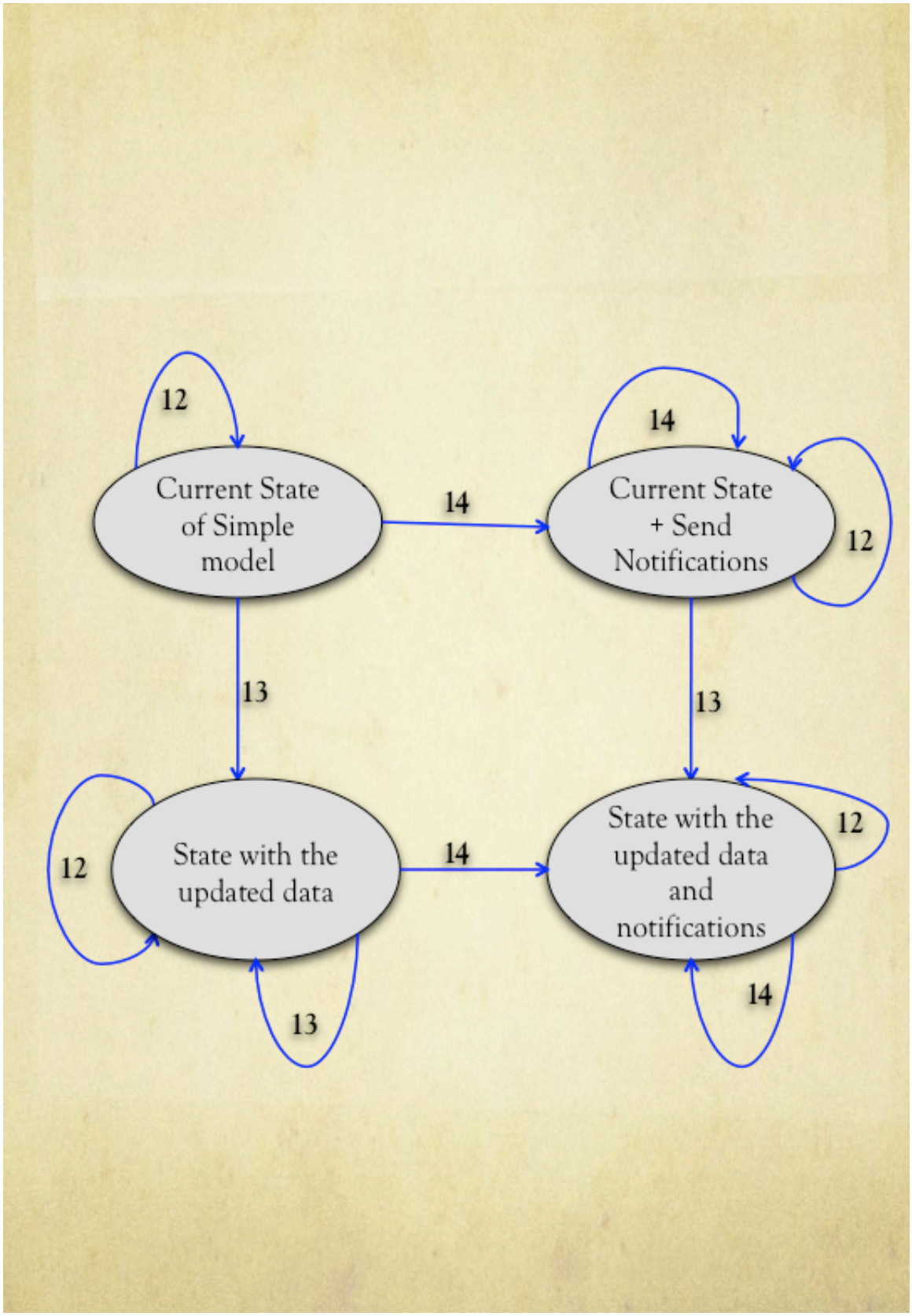


Figure 4.8: The state machine model depicting the "notify, update actions and the Φ conditions" examined in Maude

4.3.1 Healthiness conditions relating to the characteristics of the system

There are three different kinds of healthiness conditions that relate to the characteristics of the system. These are:

1. Conditions that are axiomatic for the logic.
   (a) The initial state of the system for a specific user of the system contains the rights $aO\delta$ and $bR^{\dagger}\Phi\delta$ for particular actors a, b and particular Φ conditions.
   (b) Rights change from true to false and vice versa only with actions.
   (c) The Φ conditions must be consistent. For example if a data subject gives consent to a data controller to share data with a specific third party, there cannot exist a variable in the same Φ that constrains the action denoting not to share the data with the same third party. Not all the variables of the Φ constraints are defined in the thesis, thus a proof that these must be consistent cannot be presented.
2. Desirable properties that derive from the axioms of the logic.
   (a) The sequence of the actions that do not refer to the same data is commutative (Action A(δ) ; Action B(δ') = Action B(δ') ; Action A(δ)). This condition holds because the pre-conditions of these actions are disjoint and the post-conditions are compatible (i.e. do not interact).
   (b) Actions that refer to different data items are independent. This condition holds because the pre-conditions of these actions are disjoint and the post-conditions are compatible (i.e. do not interact).
   (c) $bS^{*}\delta \implies bS\delta \implies bP\delta \implies bL\delta$.
   (d) When data is shared with third parties, at least the restrictions described in Φ, are imposed on third parties as well.

4.3.2 Healthiness conditions relating to the actors of the system

1. Data subjects can always delete (theoretically) all of their data, except where law dictates otherwise. For any data item, a state must be reachable, in which no data is stored by data controllers or third parties.

2. Third parties cannot revoke rights from data controllers.

3. Data subjects can request revocation of rights and deletion of data from the third parties, without requesting revocation of rights or deletion of data from the data controllers.

4. Data subjects can revoke rights and delete data from the data controllers, without revoking rights or deleting data from the third parties. But they still have control over their data stored in third parties.

5. Data controllers can store data without exercising any process or sharing due to law obligations (for example Internet Service Providers).

6. Data subjects' updates cannot influence the rights that a data controller has over that data.

7. Data subjects cannot impose controls on data that derived from processing or aggregation of their data.

4.4 Proofs for healthiness conditions in Maude

This section presents the proofs of the healthiness conditions described in Section 4.3. The state-model is tested with the application of counterexamples to identify undesired behaviour, to reveal possible forbidden transitions in the system and to provide proofs that every transition described by the formalised actions is achieved. The code developed for the Maude software can be found in Appendix A.

4.4.1 Right to share implies right to process

The fifth healthiness condition in Section 4.3.1 denotes that $bS^*\delta \implies bS\delta \implies bP\delta \implies bL\delta$, a condition that is also implied by the inner circles in Figure 3.1. Figure 4.9 illustrates the results obtained from Maude. The command "search a => M:State" seeks all the transitions from the initial state to such a state where $bS^*\delta \implies bS\delta \implies bP\delta \implies bL\delta$ is true. This is achieved with the use of an invariant that when true the healthiness condition holds.

In addition I sought a counterexample which requested Maude to reach a state where the condition $bP\delta \implies bL\delta$ is not true. The results are presented in Figure 4.10. From the results one can safely assume that there does not exist a state

```
Maude> search a =>+ M such that rights( M ) implies ( P and S ) .
search in SIMPLE-4 : a =>+ M such that rights(M) implies P and S = true .

Solution 1 (state 2)
states: 3  rewrites: 12 in 0ms cpu (0ms real) (96774 rewrites/second)
M --> c

Solution 2 (state 3)
states: 4  rewrites: 19 in 0ms cpu (0ms real) (98958 rewrites/second)
M --> d

Solution 3 (state 5)
states: 6  rewrites: 41 in 0ms cpu (0ms real) (133116 rewrites/second)
M --> f

Solution 4 (state 6)
states: 7  rewrites: 55 in 0ms cpu (0ms real) (142118 rewrites/second)
M --> g

No more solutions.
states: 10  rewrites: 108 in 0ms cpu (0ms real) (183050 rewrites/second)
```

Figure 4.9: States where sharing implies processing is true

where the condition $bS\delta \implies bP\delta$ does not hold, while there are four states where the condition holds, verifying the holding of this healthiness condition.

```
Maude> search a =>+ M such that rights( M ) implies (P and (not L)) .
search in SIMPLE-4 : a =>+ M such that rights(M) implies P and not L = true .

No solution.
states: 10  rewrites: 164 in 0ms cpu (0ms real) (278438 rewrites/second)
```

Figure 4.10: States where processing implies storing data is not true

4.4.2 Deletion of data from every state of the system

The first healthiness condition presented in Section 4.3.2 stated that "data subjects can always delete (theoretically) all of their data, except where law dictates otherwise. For any data item, a state must be reached, in which no data is constrained by rights". In order to prove this condition, the transitions from each state of the system to the initial state were tested in Maude. The results are illustrated in Figure 4.11 and in Figure 4.12.

```
Maude> search a =>+ b .
search in SIMPLE : a =>+ b .

Solution 1 (state 1)
states: 2  rewrites: 1 in 0ms cpu (0ms real) (13333 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (0ms real) (209912 rewrites/second)
Maude> search b =>+ a .
search in SIMPLE : b =>+ a .

Solution 1 (state 3)
states: 4  rewrites: 4 in 0ms cpu (0ms real) (63492 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (0ms real) (296296 rewrites/second)
Maude> search c =>+ a .
search in SIMPLE : c =>+ a .

Solution 1 (state 3)
states: 4  rewrites: 5 in 0ms cpu (0ms real) (87719 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (0ms real) (243243 rewrites/second)
Maude> search d =>+ a .
search in SIMPLE : d =>+ a .

Solution 1 (state 2)
states: 3  rewrites: 5 in 0ms cpu (0ms real) (70422 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (0ms real) (221538 rewrites/second)
Maude> search e =>+ a .
search in SIMPLE : e =>+ a .

Solution 1 (state 8)
states: 9  rewrites: 35 in 0ms cpu (0ms real) (248226 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (0ms real) (258064 rewrites/second)
Maude> search f =>+ a .
search in SIMPLE : f =>+ a .

Solution 1 (state 7)
states: 8  rewrites: 26 in 0ms cpu (0ms real) (189781 rewrites/second)
empty substitution
```

Figure 4.11: The results that Maude produced when the transitions from all states to the initial were tested

```
Solution 1 (state 7)
states: 8  rewrites: 26 in 0ms cpu (0ms real) (189781 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (9ms real) (200557 rewrites/second)
Maude> search g =>+ a .
search in SIMPLE : g =>+ a .

Solution 1 (state 6)
states: 7  rewrites: 16 in 0ms cpu (0ms real) (84210 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (0ms real) (164383 rewrites/second)
Maude> search h =>+ a .
search in SIMPLE : h =>+ a .

Solution 1 (state 9)
states: 10  rewrites: 42 in 0ms cpu (0ms real) (224598 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (1ms real) (215568 rewrites/second)
Maude> search i =>+ a .
search in SIMPLE : i =>+ a .

Solution 1 (state 4)
states: 5  rewrites: 5 in 0ms cpu (0ms real) (78125 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (0ms real) (264705 rewrites/second)
Maude> search j =>+ a .
search in SIMPLE : j =>+ a .

Solution 1 (state 4)
states: 5  rewrites: 7 in 0ms cpu (1ms real) (35000 rewrites/second)
empty substitution

No more solutions.
states: 10  rewrites: 72 in 0ms cpu (1ms real) (161434 rewrites/second)
Maude>
```

Figure 4.12: The results that Maude produced when the transitions from all states to the initial were tested

It is clear from the results that one can successfully transit from any state of the system to the initial state, where no rights pertain to the data, thus there does not exist a situation where personal data can be handled without the data subject being able to delete it.

4.4.3 Third parties cannot revoke rights from data controllers

The second healthiness condition in Section 4.3.2 requires that third parties are not allowed to revoke data from data controllers. In order to prove this condition, I searched for a counterexample where there would exist a state such that the third party would possess at least one of the rights to store, process, share or share transitively data (the T right in Maude code) while the data controller would have no knowledge of that possession (the X right in Maude code). The results are presented in Figure 4.13 and such a state does not exist.

```
Maude> search a =>+ M such that rights( M ) implies ( T  and (not X )) .
search in SIMPLE-4 : a =>+ M such that rights(M) implies T and not X = true .

No solution.
states: 10  rewrites: 154 in 0ms cpu (4ms real) (334056 rewrites/second)
```

Figure 4.13: Requesting to reach a state where the third party would possess rights but the data controller would not

4.4.4 Revoking rights from third parties without influencing the rights of the data controller

The third healthiness condition in Section 4.3.2 stated that "data subjects can request revocation of rights and deletion of data from the third parties, without requesting revocation of rights or deletion of data from the data controllers". In order to prove that such a condition is possible, a transition from a state where the data controller possess the rights L, P, S and the third party at least one of the L, P, S rights to a state where the rights of the data controller will remain untouched while the third party will possess no rights at all, is sought. The results are presented in Figure 4.14 and such a transition is achievable.

```
Maude> search e =>+ M such that rights(M) = O and L and P and S and S1  .
search in SIMPLE-4 : e =>+ M such that rights(M) = O and L and P and S and S1 .

Solution 1 (state 7)
states: 8  rewrites: 31 in 0ms cpu (0ms real) (196202 rewrites/second)
M --> d

No more solutions.
states: 10  rewrites: 74 in 0ms cpu (0ms real) (233438 rewrites/second)
Maude>
```

Figure 4.14: Revoking rights from third parties without influencing the rights of the data controller

4.4.5 Revoking rights from a data controller without influencing the rights possessed by third parties

The fourth healthiness condition in Section 4.3.2 stated that "data subjects can request revocation of rights and deletion of data from data controllers, without requesting revocation of rights or deletion of data from parties". In order to prove that such a condition is possible, a transition from a state where the data controller possess the rights L, P, S and the third party at least one of the L, P, S rights to a state where the rights of the data controller are revoked while third parties still possess rights, is sought. The results are presented in Figure 4.15 and such a transition is achievable.

```
Maude> search e =>+ M such that rights(M) = O and T and X  .
search in SIMPLE-4 : e =>+ M such that rights(M) = O and T and X .

Solution 1 (state 3)
states: 4  rewrites: 8 in 0ms cpu (4ms real) (49079 rewrites/second)
M --> i

No more solutions.
states: 10  rewrites: 74 in 0ms cpu (5ms real) (160520 rewrites/second)
```

Figure 4.15: Revoking rights from data controller without influencing the rights of third parties

4.4.6 Store data without being able to process

The fifth healthiness condition presented in Section 4.3.2 required that data controllers can store data even without being able to process it. A transition from the initial state to a state where the data controller will possess only the right L

is requested to prove the condition and the results are presented in Figure 4.16, illustrating that such transition is possible.

```
Maude> search e =>+ M such that rights(M) = O and L .
search in SIMPLE-4 : e =>+ M such that rights(M) = O and L .

Solution 1 (state 9)
states: 10  rewrites: 45 in 0ms cpu (0ms real) (252808 rewrites/second)
M --> j

No more solutions.
states: 10  rewrites: 74 in 0ms cpu (0ms real) (268115 rewrites/second)
Maude>
```

Figure 4.16: Reaching a state where the data subject can only store data

4.4.7 All the achievable states and the rights that pertain to them

To conclude with the Maude proofs, I present all the states that can be reached from the initial one and the rights that pertain to these. The results are presented in Figure 4.17 and there does not exist a state which has not been considered by the logic.

```
Maude> search a =>+ M .
search in SIMPLE-4 : a =>+ M .

Solution 1 (state 1)
states: 2  rewrites: 1 in 0ms cpu (5ms real) (9259 rewrites/second)
M --> b

Solution 2 (state 2)
states: 3  rewrites: 2 in 0ms cpu (6ms real) (8000 rewrites/second)
M --> c

Solution 3 (state 3)
states: 4  rewrites: 3 in 0ms cpu (6ms real) (8450 rewrites/second)
M --> d

Solution 4 (state 0)
states: 4  rewrites: 7 in 0ms cpu (6ms real) (15053 rewrites/second)
M --> a

Solution 5 (state 4)
states: 5  rewrites: 8 in 0ms cpu (6ms real) (14545 rewrites/second)
M --> j

Solution 6 (state 5)
states: 6  rewrites: 13 in 0ms cpu (7ms real) (19461 rewrites/second)
M --> f

Solution 7 (state 6)
states: 7  rewrites: 21 in 0ms cpu (7ms real) (27166 rewrites/second)
M --> g

Solution 8 (state 7)
states: 8  rewrites: 34 in 0ms cpu (7ms real) (37777 rewrites/second)
M --> i

Solution 9 (state 8)
states: 9  rewrites: 35 in 0ms cpu (7ms real) (35641 rewrites/second)
M --> h

Solution 10 (state 9)
states: 10  rewrites: 37 in 1ms cpu (8ms real) (34322 rewrites/second)
M --> e

No more solutions.
states: 10  rewrites: 64 in 1ms cpu (8ms real) (52588 rewrites/second)
```

Figure 4.17: Achievable states from the initial state and the rights that pertain to them

4.5 Lessons learnt and implications for EnCoRe

The process of designing the logic was dynamic and the final version presented in the thesis was reached by refining the logic to address issues and ambiguities that emerged through out the life-cycle of the thesis. This process of refining the logic also resulted in presenting solutions at an operational level for the EnCoRe project and the lessons learnt would have remained undiscovered without the logic. The main refinement occurred as a response to the ambiguities that emerged when the requirements of the pilot case study were formalised.

4.5.1 Ambiguities

The ambiguities that emerged from the formalisation of the requirements can be categorised into two classes. The first class comprises the ambiguities created from the application of the details of law, regulation, policy and social factors by computer scientists [259]. The second class consists of ambiguities emerge from the complexity of the notion of privacy and the gap that exists between the data subject's demand to control the flow of their data and the data controller's desire to reduce data subjects' interference. The detection and resolution of these ambiguities requires the use of formal methods.

Ambiguities of the first kind

In the first class ambiguities occur when the data subject performs actions to update, delete, revoke or change his or her given consent. More specifically, in the case where the data subject wishes to update their personal data, there can be ambiguities emerging as to whether previous data should be deleted or linked with the new data. Furthermore, in the case where the organisation has shared this data transitively it is not clear whether the changes should affect the third parties as well.

Revocation of consent, even though it is analysed in detail in the model, obtains different meanings depending on the circumstances and purposes that the data is being held for. In the case of deletion, ambiguities emerge from differences on how higher level people perceive the notion of deletion and how it could be technically implemented. For example deleting data could have multiple meanings. One could

render the data useless, scramble data, delete it from the back-up system or physically destroy the hard discs.

Ambiguities of the second kind

The second class of ambiguities highlights the complexity of the privacy problem and underlines the conflicts that emerge between a data subject and a data controller. The most interesting issues, but at the same time most difficult to address, are that of aggregation and anonymity. The complex nature of these issues could lead to a situation where it may be technically infeasible to express data subjects' preferences or the privacy regulations in place.

Aggregation unveils more information about the data subject, by combining it with information already available. Data could be processed and shared in the proper way by the data controller, but when aggregated, this data could create new information that can compromise data subject's privacy. The logic is designed in such a manner that every time data is shared it can be aggregated by the data controller, meaning that as long as they have permission to store or process data, they have the potential to aggregate that data. A possible solution to the problem of aggregation is for the data subject to define the purpose for which the data is shared and also control what further personal data the data controller collects.

The ambiguities that arise in the case of anonymity concern the way which data is anonymised. These ambiguities where unveiled when I tried to formalise a use case where the data subject requested her medical records to be anonymised if shared with another third party. Although the data subject may consent to share the anonymised medical records, the danger of the identity to be unveiled always lurks, as "data can either be useful or perfectly anonymous but never both" [195]. "There is growing evidence that data anonymisation is not a reliable technical measure to protect privacy. The development of more powerful re-linking technology capabilities and the wider access to increasing amounts of data from which to drive these are reducing its effectiveness" [84].

Even if methods such as k-anonymity [195] become efficient, the link between the data controller and the third party captured by the logic can lead to de-anonymisation of the data. The logic can capture data subjects' requests to anonymise data first and then disseminate to a data controller if the anonymised data is treated as new data. If this is the case, the data subject has no controls on that data and further sharing of the old data between the data subject, the data controller and

the third parties that have access to the anonymised data should be forbidden.

Difficulties also emerge when the data subject exercises their right to revoke consent but at the same time the data controller is unable to perform such an action. For example a data subject may request his data to be deleted but the organisation is still processing his data. To address these issues a variable allowing the data subjects to express transparency in their decisions is required. Additionally, the conflicts that occur between the data subjects and the data controllers are addressed by introducing a combination of permissions and obligations under certain conditions. For example, a data subject's request regarding the revocation of consent to process data is executed under the condition that the data controller does not process the specific data at that time.

Further ambiguities occur with the handling of meta-data. In the logic, meta-data mainly comprises of variables and ranges of values that set the context and describe data subject's consent and revocation preferences. In the conceptual models it is not clear what happens with the data subject's control preferences. An interesting example is that of notification. How can an enterprise notify a data subject that their data was deleted completely if they do not keep their email and the consent and revocation preferences describing the conditions for the action of notification?

4.5.2 Refinement process

Tackling the ambiguities created both from the formalisation of law, regulation and social factors and from the complexity of the notion of privacy, enhanced the effectiveness of the logic by introducing new actions and rights that enriched its descriptiveness. Thus, more options are can be captured and offered to the data subject to express their preferences.

The refinement of the logic comprised of novelties that allowed the effective and unambiguous formalisation of all the requirements not only for the pilot case study but for those described in Chapter 6 and Chapter 7. More specifically, in relation to the conceptual model presented in Chapter 3, four actions are introduced for updating data, enabling data subjects to update data either by: deleting the previous data; by linking data with the new; or by propagating the updates to third parties as well. In addition an action for notification is defined and an obligation is created for the data controller providing the means to the data subjects to be notified under certain conditions and through certain communication channels (e-mail etc.).

Further to the introduced actions, rights that determine whether the actions will be completed are defined. The data subject now has the right to be notified, the right to update data and the right to delegate rights to other individuals. Furthermore, the data controller has the right to know the location of every meta-data, enabling the data subject to express preferences on the way that the meta-data will be treated. Last but not least, the effectiveness of the actions is increased by the variables defined in different contexts. Variables can determine when the data subject is allowed to revoke permissions from the data controller, the way to delete data and the purpose for which the data will be used in order to address the problem of aggregation. Figure 4.18 illustrates why and how the extended Hoare logic presented here is a refinement of the simple Hoare logic.

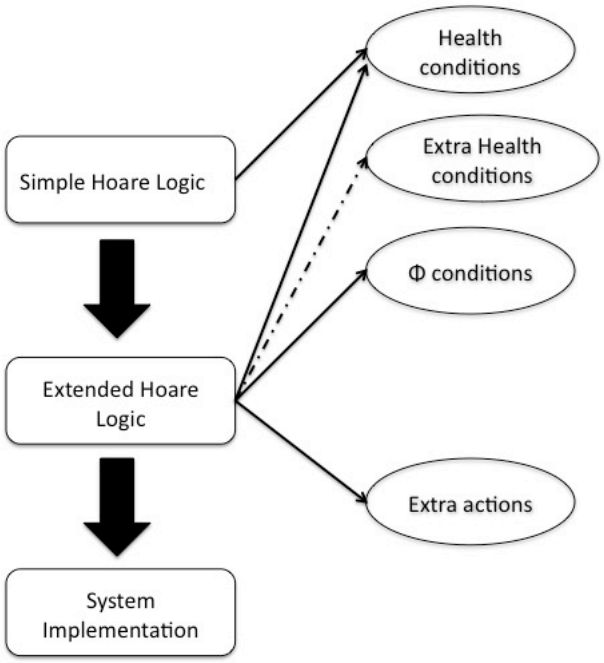


Figure 4.18: Refinement of the logic

4.6 Synopsis of the findings presented in the chapter

In this chapter I have presented a novel Hoare-style logic of consent and revocation. This is the second, and most significant, step following the conceptual model, towards a system capable of providing data subjects with consent and revocation controls. The logic is given in the form of triples comprising of pre conditions, triggered actions and post conditions. Since, to my knowledge, there does not exist a logic that addresses the issues of consent and revocation in the current literature, the logic is the main contribution of this thesis to knowledge. The core elements of the logic are rights that principals of the system possess under certain circumstances; actions that describe transitions between states of the system; obligations that capture the necessity of future actions (consequent on the one just performed); and Φ constraints that pertain to data and comprise of variables that define further parameters of consent and revocation preferences. All the rights and actions are presented and the formalisations of the desirable transitions in the system are provided.

Maude was used to check that several important properties of the logic, such as evidence that no more transitions could be achieved in the system or no undesired behaviour is expected, always hold. I achieved this by exploring all possible reachable states. These desired properties formed the healthiness conditions that the logic satisfies and verify the correctness of the mathematics. In addition, because the Hoare-style favours sequential actions, a concurrent model was illustrated, to identify problems that emerge when actions run in concurrent manner and solutions were provided to the acknowledged issues.

The logic was underpinned by the elements dictating the conceptual model and the initial attempt to specify the requirements of the pilot case study resulted in ambiguities on the formalisation. These ambiguities were addressed and a refined version of the logic occurred. The designing process was a gradual refinement of the logic and the version presented in this Chapter is the final version, applied to all three case studies of the EnCoRe project. The chapter concludes with the addressed ambiguities and the lessons learnt from the formalisation, that otherwise would have remained veiled.

CHAPTER 5

The Employee case study

Having designed a language able to capture the effects of data subjects expressing consent and revocation preferences and having proven with the use of Maude the healthiness conditions satisfied by that language, the next step is to verify its richness and efficiency when applied to diverse contexts. As shown in Chapter 3, data subjects have different privacy expectations, which depend upon the power relations developed when they initiate a relationship with a data controller and the type of data they disclose. The logic must adequately formalise the different requirements elicited from diverse environments.

In order to verify the expressiveness of the logic, I adopted the research methodology of case studies. The case studies presented in this thesis were designed by the EnCoRe consortium for the purposes of the EnCoRe project. I was fortunate to acquire the benefits of using the EnCoRe case studies, because they were inspired by real world scenarios and driven by business needs. The names of the enterprises that the case studies were designed to serve will not be revealed, due to confidentiality issues.

Orlikowski et al [196] argue that the case study methodology is the most popular approach to be followed in the field of IS because it is "an empirical inquiry that investigates a contemporary phenomenon within its real life context" [272]. My decision to apply the logic on the EnCoRe case studies was informed by the fact that there are three different power relations created between a data controller and a data subject. Each of the three case studies of the EnCoRe project is designed to capture one of the three different types of relations that craft the data subjects' consent and revocation expectations.

In this Chapter, I present the formalisation of the pilot case study, named the Employee case study. The choice for the first EnCoRe case study was informed by

the fact that the management of employee data in organisations is a well-understood problem, and employees' privacy offers interesting issues in terms of managing consent and revocation controls in a context where different business and legal requirements need to be taken into account.

The case study describes a number of use case scenarios and elicits from these a list of requirements. The implications of invoking consent and revocation controls are explored. These use cases are meant to illustrate key points affecting the management of consent and revocation, such as the provision or revocation of consent by a data subject; enforcing consent and revocation preferences; dealing with the overall consent and revocation controls and its impact on data including notifications, updates, etc.

In the employee data scenario, the focus is on types of personal data, such as trade union membership, financial/payment detail, home address details, family details, etc. Personal data can be gathered by different sub-organisations within the enterprise (e.g. HR department, Payroll, Occupational Health department etc.) or can be shared with third parties (web services to sell products, etc.). The concept of collaboration among employees, referred as Workbook, is introduced to the case study to capture the situation where a data subject interacts and shares data with another data subject. Although this scenario is far from what can be achieved today in terms of consent and revocation controls, I believe that the emerging requirements, with reality checks, contribute to the understanding of these controls and could indicate problems and ambiguities when implementing them.

In Section 5.1 bellow, the methodology followed to elicit the requirements for capturing consent and revocation preferences is presented. In Section 5.2, the use cases and their formalisation are discussed. In addition, the ambiguities which emerged from applying the first version of the logic are mentioned and the refined actions and rights that addressed them are illustrated. I will also give a few examples of what values the variables may take and how data subjects' choices could be changed or revoked. Section 5.3 provides a synopsis of the results presented in this Chapter.

5.1 Eliciting requirements

Requirements usually describe the necessary or expected behaviour of the system in place, or the constraints to behaviours which would be observed in the system. The

preferred approach for eliciting requirements in this thesis was to study the relevant academic literature, to analyse the legacy systems already in place and to gain an in-depth understanding by analysing the results of user requirements workshops, specifically designed for the needs of the EnCoRe project. However, it was not feasible to undertake all these steps to elicit requirements for all the case studies and different approaches were adopted.

The focus groups for this case study were generic and not designed for the specific scenario, unlike the focus groups for the Biobank case study. For this reason, the Goal Structuring Notation (GSN) [145] was used to elicit and analyse the main requirements for a system managing consent and revocation preferences. This notation is useful for expressing safety-related requirements and it provides means for describing the desired functionality of a system. The six steps involved in the development of a goal structure are:

1. Identify goals to be supported;
2. Define basis on which goals are stated;
3. Identify strategy to support goals;
4. Define basis on which strategy is stated;
5. Elaborate strategy (and therefore proceed to identify new goals - repeat the process for the new goals);
6. Identify the basic solution.

The requirements are expressed in terms of the different types of information needed at each level of the system and the GSN is used to break down overall system goals. Firstly, I identify which standard messages would be exchanged in a consent and revocation management system and models of how, and in which sequence these messages are transmitted, are built.

5.2 Formalising the Employee case study

The description of the use cases, which are independent of each other, is captured in italics and the symbols used in the following formalisations are:

- m denotes Mary.

- b denotes her boss
- h denotes the enterprise's Human Resources department.
- z denotes Phil
- δ, δ_1, δ_2 represent data.
- c denotes a third party with whom the enterprise shares data.
- c_1 is the company that Mary shares data with when she is abroad.
- y is the Company for which Phil works.
- k is the next of kin.
- s is a variable to denote that Mary is sick.
- q is a variable to denote the reason for the delegation process to take place.

5.2.1 Mary is hired

Consent for processing and sharing one step further

Mary is hired at Company X. She has to go through the hiring process and provide personal data: address, financial details, etc. Before starting to work, she fills out forms for the Human Resources (HR) department. These reports require the disclosure of health information, next of kin references, etc. She signs a form agreeing to these terms and conditions which are then stored by HR.

$$\{mO\delta \wedge hR^{\dagger}\Phi\delta\}$$
$$\mathbf{grant^1}(m, h, \Phi, \delta)$$
$$\{hL\delta \wedge hP\delta \wedge hS\delta \wedge hR\Phi\delta\}$$

where Φ = parties to access data:Π^* and $\Pi^* \subseteq \{$HR department, Boss$\}$.

In this formalisation, Mary discloses δ to the HR department. She also gives consent to the HR department to process her data and share it only one step further. There is no further information on what data Mary discloses to the HR department. If this information was available, each data could have been set as $\delta_1, \delta_2, \delta_3$ and

controls could pertain to each data separately. Data can also be classified into larger groups and controls can be posed over a set of data. However, the issue of scalability should be considered because it may not be feasible for real-world systems to cope with controls in every single datum stored. There are implications both for the data subject (how could they be efficiently informed on how to control each piece of data) and for the data controller (is it cost effective or even technically feasible to provide such scalable controls? Are there any complications or delays for the business operation?).

Note in the formalisation the constraint regarding the processing of data, as only the HR department and Mary's boss may obtain access to process it. As formalised in later use cases, Mary could set additional Φ conditions pertaining to the processing or sharing of data and the notification procedure.

HR shares Mary's data with her boss

Some of the information, which Mary provided to the HR department during her interview (CV, presentations, application and interview forms, etc.), might be relevant to Mary's boss.

With the action below the HR department provides to Mary's boss access to her personal data. The previous use case formalised Mary's choice to allow her boss to access her personal data. In the above formalisation, since the Φ conditions are respected, Mary's boss b obtains access to Mary's data and process it.

$$\{hS\delta \wedge bR^{\dagger}\Phi\delta \wedge b \in \cup\Phi.\text{parties to access data}\}$$
$$\mathbf{grant}(h, b, \Phi, \delta)$$
$$\{bL\delta \wedge bP\delta \wedge hXb\delta \wedge bR\Phi\delta\}$$

This formalisation denotes that the HR department can only share Mary's data δ when in the pre-condition the HR department has the right to share the data δ. Mary has expressed her preference to allow her boss to obtain the data and this is captured with the $b \in \cup\Phi$.parties to access data notation. Maybe the HR could impose some default controls on Mary's boss since Mary has not so far expressed any other preferences. I will demonstrate how this formalization could be applied in a following example.

The processing of data from Mary's boss is captured in the formalisation below:

$$\{bR\Phi\delta \wedge bP\delta\}$$
$$\textbf{process}(b, \Phi, \delta)$$
$$\{true\}$$

The post-condition denotes that the action of processing has taken place and since Φ conditions do not exist, the restriction in Φ that may happen, as demonstrated in Chapter 4, does not apply in this scenario.

Mary allows the HR department to share her data with third parties

Mary gives consent to other services to process and share transitively her data only if they notify her. She is asked to join a few mandatory enterprise services, such as the employee portal, pension scheme, travel service, etc. She is offered the opportunity to express notifications preferences about access/internal usage/disclosure to third parties of the data.

$$\{mO\delta \wedge mN^{\dagger}\delta \wedge hP\delta \wedge hR\Phi\delta \wedge hR^{\dagger}\Phi^{*}\delta\}$$
$$\textbf{setnotify}(m, h, \Phi^{*}, \delta)$$
$$\{hLn^{\dagger} \wedge hLn^{*} \wedge hR\Phi^{*}\delta \wedge hNa\delta\}$$

where Φ = notify-how:$n^{*} \wedge$ notify-what:$n^{\dagger}$ and $n^{*} \subseteq$ {email} and $n^{\dagger} \subseteq$ {shared data with third parties}.

In this formalisation Mary sets her notification choices and she selects to be notified via email when the HR department shares her data.

$$\{hS\delta \wedge hR\Phi\delta \wedge cR^{\dagger}\Phi^{*}\delta \wedge c \in \cup\Phi.\text{destination} \wedge \Phi^{*} \leq \Phi\}$$
$$\textbf{grant}(h, c, \Phi^{*}, \delta)$$
$$\{cL\delta \wedge cP\delta \wedge hXc\delta \wedge cR\Phi^{*}\delta \wedge$$
$$\forall c.\langle hNa\delta \wedge \text{“shared”} \in \cup\Phi.\text{notify-what} \wedge \text{“email”} \in \cup\Phi.\text{notify-how}$$
$$\rangle\ \textbf{notify}(h, m, \delta, \text{“shared”}, \text{“email”})\ \langle true\rangle\}$$

where Φ = purpose:u

$\wedge$ time duration:$t \wedge$ times processed:t^* and $u \subseteq$ {internal usage}, t= 5 months and t^*= 100 times processed and $c \subseteq$ {the employee portal, pension scheme, travel service}.

In addition, the HR department shares her data with third parties and imposes controls regarding the usage of data. There is an obligation included in the post-condition, which is triggered only if Mary has set notification requests. Note that the sequence in which the two actions will be executed in irrelevant, since the final result is the same in both possible combinations.

5.2.2 Mary is enrolled in other services

Mary allows third parties to process her data and sets notification restrictions

Mary decides to voluntarily join a few initiatives offered by the Company, including a HW&SW purchase schema (offered by an external company), Sport and Social Club (SSC) and Holiday Cottage Service - all provided by third party companies. Mary's personal data, including address, employee references and financial details, has to be disclosed to these companies. She is also required to express her notification preferences.

$$\{hS\delta \wedge hR\Phi\delta \wedge cR^{\dagger}\Phi^{*}\delta\}$$
$$\mathbf{grant}^{\dagger}(m, h, c, \delta, \Phi^{*})$$
$$\{cL\delta \wedge cP\delta \wedge cR\Phi^{*}\delta \wedge$$
$$\langle mO\delta \wedge mN^{\dagger}\delta \wedge hR^{\dagger}\Phi'\delta\rangle\ \mathbf{setnotify}(m, h, \Phi', \delta)\ \langle hNa\delta \wedge hR\Phi'\delta\rangle \wedge$$
$$\forall c.\langle hNa\delta \wedge mN^{\dagger}\delta \wedge \text{“shared”} \in \cup\Phi'.\text{notify-what}$$
$$\wedge \text{“email”} \in \cup\Phi'.\text{notify-how}\rangle\ \mathbf{notify}(b, a, \delta, \text{“shared”}, \text{“email”})\ \langle true\rangle\}$$

where Φ^* = purpose : u and $u \subseteq$ {internal use} and Φ' = notify-how:$n^* \wedge$notify-what:$n^{\dagger}$, $n^* \subseteq$ {email}, $n^{\dagger} \subseteq$ {shared data with third parties} and $c \subseteq$ {HW&SW purchase schema, SSC, Holiday Cottage Service}.

In this use case, a similar formalisation with the previous use case is presented. The subtle difference is that Mary requires from the HR department to disclose her data to specific third parties. If Mary has not set her notification preferences, there is an obligation for the data controller to request from Mary her notification

options. Once these options are in place, an obligation to notify Mary is triggered when the data is shared with the third parties.

5.2.3 Data outsourced

Mary is offered the opportunity to express degrees of consent about some of the personal data affected by these activities. This might include notifications about access/internal usage/disclosure to third parties of this data. Otherwise default settings might apply. Mary decides to withdraw from the voluntary SSC service. She revokes her consent to use her personal data and expresses the preference that her personal data should be deleted.

The actions of expressing degrees of consent or setting notification preferences can be formalised as in previous use cases. Below is formalised Mary's decision to revoke consent and delete data from the Sport and Social Club.

$$\{mO\delta \wedge hR\Phi\delta \wedge bXc\delta\}$$
$$\mathbf{delete}^{\dagger}(m, h, \delta)$$
$$\{\forall c.\langle hXc\delta \wedge cL\delta\rangle\ \mathbf{delete}^{\dagger}(b, c, \delta)\ \langle \neg cL\delta \wedge \neg cP\delta \wedge \neg cS\delta \wedge \neg cR\Phi\delta \wedge \neg bXc\delta\rangle\}$$

In this formalisation Mary requests her data to be deleted by all third parties. The logic can also capture situations where Mary defines from which third party in particular her permissions will be revoked. This example is formalised in use case 5.2.10.

Ambiguities in deletion

Considering deletion, ambiguities emerge from the diversity with which people perceive this issue. Further complexity occurs from the different ways which deletion can be technically implemented. For example, the multiple meanings that deleting data may have is: data can be rendered useless; data can be scrambled; data can be deleted from the back-up system; request the hard discs to be physically destroyed. This type of ambiguity is addressed in the logic with the use of the x variable,

defining the specific technical implementation for deletion.

5.2.4 Mary withdraws from some services

Mary revokes permission to process her data from services and requires her data to be deleted. Over time, She reassesses her affiliation to a few voluntary services which had joined in the past. She decides to leave a few of them, including the stock purchase program but not the pension fund service. Thus, she requests the revocation of consent regarding the use of her personal data by these services and requires her data to be deleted where possible.

The formalisation is similar to that of the previous use case. The only difference is that the pension scheme is not included in the c set.

$$\{mO\delta \wedge hR\Phi\delta \wedge bXc\delta\}$$
$$\mathbf{delete}^{\dagger}(m, h, \delta)$$
$$\{\langle hXc\delta \wedge cL\delta\rangle\ \mathbf{delete}^{\dagger}(b, c, \delta)\ \langle \neg cL\delta \wedge \neg cP\delta \wedge \neg cS\delta \wedge \neg cR\Phi\delta \wedge \neg bXc\delta\rangle\}$$

where $c \subseteq \{\text{stock purchase program}\}$.

Mary changes her notification preferences

Mary decides to change her notification preferences regarding the Holiday service, whilst still using this service.

$$\{mO\delta \wedge hR\Phi\delta \wedge hR^{\dagger}\Phi^{*}\delta\}$$
$$\mathbf{change}(m, h, \Phi^{*}, \delta)$$
$$\{\neg hR\Phi\delta \wedge hR\Phi^{*}\delta\}$$

where $\Phi^{*} = \text{notify-how:}n^{*} \wedge \text{notify-what:}n^{\dagger}$ and $n^{*} \subseteq \{\text{telephone}\}$ and $n^{\dagger} \subseteq \{\text{share}\}$.

In the formalisation presented above, Mary alters her notification preferences and requests to be notified with a phone-call every time her data is shared.

5.2.5 Mary is getting married

Mary updates her data

Mary is getting married. She has to update her personal data stored in Company X and propagate these updates to some of the third party services (including change address, financial details, indication of next of kin, etc.).

$$\{mO\delta \wedge hL\delta \wedge hR\Phi\delta \wedge hR^{\dagger}\Phi\delta' \wedge hP\delta$$
$$\wedge hS\delta \wedge bN\delta\}$$
$$\mathbf{update}(a, b, \delta, \delta')$$
$$\{bL\delta \wedge bR\Phi\delta \wedge bP\delta \wedge bS\delta \wedge bN\delta \wedge bL\delta'$$
$$\wedge bR\Phi\delta' \wedge bP\delta' \wedge \pm bS\delta' \wedge bN\delta' \wedge \forall c.\langle hXc\delta\rangle\ \mathbf{update}^{\dagger}(h, c, \delta, \delta')\ \langle true\rangle\}$$

In this formalisation, Mary updates her data but her previous address for example, is not deleted by the data controller. An obligation in the post-condition is included, in order to propagate the updated data to the services which the HR department has shared data with.

An example where previous data is deleted and the updated data is propagated to third parties

Once married, Mary must update her data but expresses the preference to delete the old data, whilst the same consent preferences should remain intact. In the case where the company X or the third parties have the right to share Mary's data onward, Mary could request her changes to be propagated to all the parties who share her data.

$$\{mO\delta \wedge hL\delta \wedge hR\Phi\delta \wedge hR^{\dagger}\Phi\delta \wedge hP\delta \wedge hS\delta$$
$$\wedge hN\delta\}$$
$$\mathbf{update}^{\dagger}(m, h, \delta, \delta')$$
$$\{\neg hL\delta \wedge \neg hR\Phi\delta \wedge \neg hP\delta \wedge \neg hS\delta \wedge \pm\neg hS^{*}\delta$$
$$\wedge \pm hN\delta \wedge hL\delta' \wedge hR\Phi\delta' \wedge hP\delta' \wedge hS\delta' \wedge \pm hS^{*}\delta'$$
$$\wedge \pm hN\delta' \wedge \forall c.\langle hXc\delta\rangle\ \mathbf{update}^{\dagger}(h, c, \delta, \delta')\ \langle true\rangle\}$$

In this formalisation, Mary updates her data but her previous address, for ex-

ample, is deleted by the data controller. The difference with the previous use case is that all the previous data is deleted.

Ambiguities occurring when updating data

When the data subject wishes to update his or her personal data, there are ambiguities emerging as to whether previous data would be deleted or linked with the new data. Furthermore, in the case where the organisation has shared their data transitively it should be clarified by the data subject whether the changes should affect the third parties as well. I decided to design the logic in such a manner that all the updates would be propagated to the third parties with whom data has been shared with. My decision is based on the fact that legislation, as depicted in Chapter 2, requires the data acquired by any data controller to be kept always up-to-date. If necessary, the logic could be reconfigured to address different priorities.

With the logic one can formalise requirements that allow data subjects to control what information is updated, how it is updated, who will process the updated information. In addition guards can be set on these controls, expressed via the values of the variables chosen by the data subject. Similar to the process of update, when an individual changes the Φ conditions that constrain the initial consent, these changes will affect the third parties. The healthiness condition requesting the Φ conditions to be consistent is crucial for this functionality.

Mary delegates consent to next of kin

Mary decides to delegate her controls over her personal data to her husband while she is abroad.

The case of delegation is formalised as follows:

$$\{mO\delta \wedge hR^{\dagger}\Phi\delta \wedge hL\delta\}$$
$$\mathbf{delegate}(m, k, \delta)$$
$$\{kU\delta\}$$

where Φ = reason for delegation:q and $q \subseteq$ {Request to delegate permissions to husband while abroad}.

In this formalisation k denotes the next of kin, who after the action is now able to act on Mary's behalf. The reason that this process took place is also captured in

the Φ conditions.

5.2.6 Mary gets promoted

Mary consents to HR to process and share additional data

Mary's internal role is changed. She is promoted and as a result she has a team to supervise. She has new duties and responsibilities. Becoming a manager implies further access to internal services (e.g. .Project Management tools, Performance Evaluation services, etc) who are going to be aware of her personal data.

$$\{mO\delta_1 \wedge hR^{\dagger}\Phi\delta_1\}$$
$$\mathbf{grant^1}(m, h, \Phi, \delta_1)$$
$$\{hL\delta_1 \wedge hP\delta_1 \wedge hS\delta_1 \wedge hR\Phi\delta_1\}$$

Φ = purpose : u and $u \subseteq$ {Project Management tools, Performance Evaluation services}.

This formalisation is similar to the first use case. Mary is giving consent to the HR department to obtain, process and share new data further. However, she allows the processing of her new data only by project management and performance evaluation tools.

Mary sets notification preferences on her new data

She is offered the opportunity to give consent about some of the personal data affected by these activities. This might include notifications about access/internal usage/disclosure of this data to third parties. Thus, notification preferences should be propagated to third parties as well. Otherwise default settings might apply.

$$\{mO\delta_1 \wedge mN^{\dagger}\delta_1 \wedge hP\delta_1$$
$$\wedge hR\Phi\delta_1 \wedge hR^{\dagger}\Phi^{*}\delta_1\}$$
$$\textbf{setnotify}(m, h, \Phi^{*}, \delta_1)$$
$$\{hLn^{\dagger} \wedge hLn^{*} \wedge hR\Phi^{*}\delta_1 \wedge hNa\delta_1$$
$$\wedge \forall c.\langle hXc\delta_1 \wedge cR^{\dagger}\Phi\delta_1\rangle \ \textbf{setnotify}(h, c, \Phi^{*}, \delta_1)\ \langle true\rangle\}$$

where Φ^{*} = notify-how:$n^{*} \wedge$ notify-what:$n^{\dagger}$ and $n^{*} \subseteq$ {email} and $n^{\dagger} \subseteq$ {share data with third parties}.

The post-condition in this permission contains an obligation which will be triggered only if the HR department has shared Mary's data. The pre-condition involves only the data controller to whom Mary consented to disclose her data and this is the reason why the action in the obligation contains the Φ conditions. In this way, these conditions will be propagated to the third parties as well.

5.2.7 Mary moves to a different country

Mary updates her data

Mary is asked to temporarily move in a new department, in a different country. Some of her data needs to be updated (temporary address).

$$\{mO\delta \wedge hL\delta \wedge hR\Phi\delta \wedge hP\delta \wedge hS\delta$$
$$\wedge hN\delta\}$$
$$\textbf{update}(m, h, \delta, \delta_2)$$
$$\{hL\delta_2 \wedge hR\Phi\delta_2 \wedge hP\delta_2 \wedge hS\delta_2 \wedge \pm hS^{*}\delta_2$$
$$\wedge \pm hN\delta_2 \wedge hL\delta \wedge hR\Phi\delta \wedge hP\delta \wedge hS\delta \wedge \pm hS^{*}\delta$$
$$\wedge \pm hN\delta \wedge \forall c.\langle hXc\delta\rangle \ \textbf{update}(h, c, \delta, \delta_2)\ \langle true\rangle\}$$

Mary updates her address by disclosing an alternative temporary address to the HR department. Her updated address is propagated to all third parties.

Mary adds new data

New information added (local bank account, health care service, etc.) as well.

$$\{mO\delta_1 \wedge hR^{\dagger}\Phi\delta_1\}$$
$$\mathbf{grant^1}(m, h, \Phi, \delta_1)$$
$$\{hL\delta_1 \wedge hP\delta_1 \wedge hS\delta_1 \wedge hR\Phi\delta_1\}$$

where Φ = purpose:$u \wedge$ time duration:$t \wedge$ times processed:t^* and $u \subseteq$ {internal use}, t= 5 months and t^*= 100 times processed.

The formalisation of this use case denotes that the new information which is available only for internal use, will be stored up to 5 months and will be processed no more than 100 times.

Mary's data is shared abroad

She also uses new local services that are provided in the new country by local service providers. Some of her personal data might need to flow across national borders.

$$\{mO\delta_1 \wedge c_1R^{\dagger}\Phi\delta_1\}$$
$$\mathbf{grant^1}(m, c_1, \Phi, \delta_1)$$
$$\{c_1L\delta_1 \wedge c_1P\delta_1 \wedge c_1S\delta_1 \wedge c_1R\Phi\delta_1\}$$

where Φ = purpose:$u \wedge$time duration:$t \wedge$times processed:t^* and $u \subseteq$ {Use for companies abroad}, t =1 month, t^* = 10 times processed and $c_1 \subseteq$ {Companies Abroad}.

When Mary sends her data abroad an assumption that the controls will be stricter is made. Thus, she decides to specify that this data is only for companies that are not based in the UK. Furthermore, the data will be stored for one month and it will be processed up to 10 times. The c_1 notation is used to denote the company abroad.

Mary sets notification preferences

She is offered the opportunity to express degrees of consent about some of the

personal data affected by these activities. This might include notifications about access/internal usage/disclosure regarding third parties which handle her data.

$$\{mO\delta \wedge mN^{\dagger}\delta \wedge c_1P\delta \wedge c_1R\Phi\delta \wedge c_1R^{\dagger}\Phi^{*}\delta\}$$
$$\mathbf{setnotify}(m, c_1, \Phi^{*}, \delta)$$
$$\{c_1Ln^{\dagger} \wedge c_1Ln^{*} \wedge c_1R\Phi^{*}\delta \wedge c_1Na\delta\}$$

where Φ^{*} = notify-how:$n^{*} \wedge$ notify-what:$n^{\dagger}$ and $n^{*} \subseteq$ {email} and $n^{\dagger} \subseteq$ {shared with third parties}.

Mary chooses to be notified by the company abroad via email, when her data is shared with third parties.

Mary revokes permissions from UK based services

She might temporarily revoke access to her personal data from services she had joined in UK, whilst she is abroad.

$$\{mO\delta \wedge hR\Phi\delta \wedge hXc\delta\}$$
$$\mathbf{revoke}^{\dagger}(m, h, \Phi, \delta)$$
$$\{\forall c.\langle bXc\delta \wedge cP\delta \wedge \pm cS\delta\rangle\ \mathbf{revoke}^{\dagger}(b, c, \delta)\ \langle cL\delta \wedge \neg cP\delta \wedge \neg \pm cS\delta\rangle\}$$

Since the data to the services was shared by the HR department, Mary requests from the HR department to get in conduct with all the services that have shared her data with and revoke permissions to process and share her data. Notice that the services will still store her data. In addition the rights of the HR department are not revoked.

Ambiguities occurring when revoking data

Ambiguities emerge when the data subject exercises their right to revoke consent but at the same time the data controller is unable to perform such an action. For example a data subject may request their data to be deleted but the organisation is still processing the data. To address these issues the use of variable p is introduced, which allows the data subjects to express transparency in their decisions. The conflicts that occur between the data subjects and the data controllers are tackled by

introducing a combination of permissions and obligations under certain conditions. For example in order for a data subject to revoke their consent to process data, there is a condition that the data controller does not process the specific data at that time (p, p^* variables).

5.2.8 Mary gets sick

Mary gets sick. She needs to stay away from her work for 6 months; she might have limited access to enterprise services. She might have expressed her preferences in terms of consent and revocation on how to handle this situation, in particular in terms of accesses to her data whilst away when she is sick.

$$\{mO\delta \wedge hR\Phi\delta \wedge hR^{\dagger}\Phi^{*}\delta\}$$
$$\textbf{change}(a, b, \Phi^{*}, \delta)$$
$$\{\neg bR\Phi\delta \wedge bR\Phi^{*}\delta\}$$

where Φ^* = destination:$\Pi \wedge \neg$forbidden destination:$\pi \wedge$ purpose:u
$\wedge$ time duration:$t \wedge$ times processed:$t^* \wedge$ notify-how:n^*
$\wedge$notify-what:$n^{\dagger} \wedge$reason for changes:r and $\Pi \subseteq$ {pension scheme}, $\pi \subseteq$ {companies abroad}, $u \subseteq$ {internal use}, t= 1 month, t^*= 10 times processed, $n^* \subseteq \{-\}$,$n^{\dagger} \subseteq \{-\}$ and $r \subseteq$ {sickness}.

In the above formalisation Mary changes her consent and revocation choices and the reason for this change is that she is sick. She decides not to be notified while she remains sick and she restricts the accessing of her data to internal use only, while prohibits the sharing of data to companies abroad.

Alternatively, this use case can be formalised by using an obligation which will be triggered when Mary is sick. In order to use obligations it would be necessary to include in the first use case, where Mary shares her data, the s variable in the Φ conditions. In the case where Mary is sick, the s variable will become true, resulting in the obligation to be triggered.

5.2.9 Team is expanding

Mary obtains rights to process and share data

Mary is expanding her team. Phil, a contractor, joins this team from company Y. Company Y has already a copy of Phil's personal data. Phil might have expressed degrees of consent and preferences about this data. Now part of this data (address, financial information, etc.) might need to be disclosed to Company X.

$$\{yR\Phi\delta \wedge yS\delta \wedge hR^{\dagger}\Phi\delta\}$$
$$\mathbf{grant^1}(y, h, \Phi, \delta)$$
$$\{hL\delta \wedge yP\delta \wedge yS\delta \wedge yR\Phi\delta\}$$

where Φ = destination:$\Pi \wedge \neg$forbidden destination:$\pi \wedge$ purpose:$u \wedge$ time duration:$t \wedge$ times processed:t^* and $\Pi \subseteq$ {pension scheme}, $\pi \subseteq$ {insurance companies}, $u \subseteq$ {internal process}, t= 5 months and t^*= 100 times processed.

In this formalisation Company Y sends Phil's data to the HR department of the Company X, along with Phil's choices.

Mary allows company X to use the new tool to process her data

Company X decides to introduce a new data analytics tool, a tool able to process emails exchanged by employees and produce graphs and information describing the flow of information and intensity. This includes linkage back to individuals (e.g. reflecting work or personal-related exchanges of emails, on work machines).In particular this applies to Mary's and Phil's email traffic.

$$\{mO\delta_2 \wedge hR^{\dagger}\Phi\delta_2\}$$
$$\mathbf{grant}(m, h, \Phi, \delta_2)$$
$$\{hL\delta_2 \wedge hP\delta_2 \wedge hR\Phi\delta_2\}$$

where Φ = purpose:$u \wedge$ time duration:$t \wedge$ times processed:t^* and $u \subseteq$ {Use for the new tool}, t= 1 month and t^*= 10 times processed.

Mary discloses personal data to the HR department, specifically to be processed only by the new tool.

Mary changes permissions for various organisations via "WorkBook"

Mary has a bad week at work and every day finds an email message offering insurance from a company associated with the "WorkBook". She notices from her daily reports (sent by "WorkBook") that her current profile is primarily viewed by "dubious" organisations (including marketing and financial sites). She is annoyed by this intrusion and as a result accesses the "WorkBook" control panel to alter her consent parameters. She sets her preferences, regarding the type of offers received, to decline offers relating to financial services.

$$\{mO\delta_2 \wedge hR\Phi\delta_2 \wedge hR^{\dagger}\Phi^{*}\delta_2\}$$
$$\textbf{change}(m, h, \Phi^{*}, \delta_2)$$
$$\{\neg hR\Phi\delta_2 \wedge hR\Phi^{*}\delta_2\}$$

where Φ^{*} = forbidden destination:$\pi \wedge$ time duration:$t \wedge$ times processed:t^{*} and $\pi \subseteq$ { Insurance companies, Finance companies }, t= 1 month and t^{*}= 10 times processed.

In this formalisation, Mary changes her consent choices and the new tool cannot share her data with insurance companies or with companies offering financial packages.

5.2.10 Termination of contracts

Mary terminates Phil's contract

Mary is notified about Phil's abnormal behaviour, both in terms of emails exchanged with external people and access to gambling sites. This is against Company X policies. Phil cannot justify the reasons of this behaviour. Mary has to terminate his contract. Phil might still be entitled to revoke consent to use his personal data from the Company X. Alternatively this could be done automatically, based on Phil's preferences.

Since Phil was an employee of Company Y, before employed by company X for a short-term position, I assume that Company X had acquired data from Company Y and not directly from Phil. Thus, the action formalised denotes that Phil requested from Company Y to get in conduct with Company X in order to revoke from them the permissions to share and process his data.

$$\{zO\delta \wedge hP\delta \wedge hS\delta \wedge hR^{\dagger}\Phi\delta \wedge yXh\delta\}$$
$$\mathbf{revoke}^{\dagger}(z, y, h, \Phi, \delta)$$
$$\{\langle yXh\delta\rangle\ \mathbf{revoke}^{\dagger}(y, h, \delta)\ \langle hL\delta \wedge \neg hP\delta \wedge \neg hS\delta \wedge yXh\delta\rangle\}$$

The formalisation describes Phil's choice to revoke permissions from the Company X to process and share his data. However, the Company X still possess his data, while the rights of the Company Y remain intact.

The situation where Phil sets an automated process to request the revocation of permissions once he stops working for Company X, can be captured with a boolean variable in the Φ conditions that once Phil is made redundant, will become false and trigger an obligation in the system similar to the action formalised above.

Mary is getting demoted

Mary is unhappy with the working environment. Her team is shrinking and she is frustrated with her inability to deliver. She is eventually demoted. She feels there has been a breach of trust. She decides that the consent she gave regarding her personal data should be revoked along with the revocation of subscription to a few corporate services (e.g. stock purchase scheme, etc.).

$$\{mO\delta \wedge hR\Phi\delta \wedge hXc\delta\}$$
$$\mathbf{revoke}^{\dagger}(m, h, \Phi, \delta)$$
$$\{\forall c.\langle bXc\delta \wedge cP\delta \wedge \pm cS\delta\rangle\ c.\mathbf{revoke}^{\dagger}(b, c, \delta)\ \langle cL\delta \wedge \neg cP\delta \wedge \neg \pm cS\delta\rangle\}$$

This formalisation is similar to the one described in the use case 5.2.7. Mary decides to revoke permissions from all the third parties.

Mary sees advertisement for skiing holidays in "WorkBook"

Mary accesses the "WorkBook" site one day and sees an advert for a group skiing holiday in the Alps. She discusses this with some employees through "WorkBook" and they agree it is an offer which can not be refused. The group decides to sign up for the skiing holidays.

$$\{mO\delta \wedge hR^{\dagger}\Phi\delta \wedge \pm bP\delta \wedge \pm bS\delta \wedge\}$$
$$\mathbf{grant}^{\dagger}(m, h, c, \Phi, \delta)$$
$$\{hP\delta \wedge hS\delta \wedge hR\Phi\delta \wedge$$
$$\langle cR^{\dagger}\Phi\delta\rangle\ \mathbf{grant}(b, c, \Phi, \delta)\ \langle cL\delta \wedge cP\delta\rangle \wedge$$
$$\langle bNa\delta\rangle\ \mathbf{setnotify}(a, c, \Phi', \delta)\ \langle true\rangle \wedge$$
$$\langle bNa\delta \wedge aN^{\dagger}\delta \wedge \Phi' \ni n^{\dagger} \ni \text{“shared”}\rangle\ \mathbf{notify}(b, a, \delta, \text{“shared”}, n^{*})\ \langle true\rangle\}$$

where $\Phi' =$ notify-how:$n^{*} \wedge$ notify-what:$n^{\dagger}$ and $n^{*} \subseteq \{\text{email}\}$,$n^{\dagger} \subseteq \{$when data is processed$\}$.

The formalisation of this use case is similar to the cases where Mary grants permissions to the HR department to process and share data. Although via the Workbook Mary has the opportunity to communicate and share information with other data subjects, resulting in a balanced power relation, the only difference in the formalisation derives by the business requirements. According to Sandhu et al, social networks require a single policy for every user, a business requirement that UCONABC as they have recognised is not able to capture yet [204]. In their position paper, they indicate their intentions to address the issue but as proven with these formalisations, the logic for consent and revocation can effectively capture unique policies for every user [204].

Mary leaves the company

Mary eventually decides to leave the company. She decides to revoke all (optional) permissions on usage of her data and requires deletion of data where possible.

Since Mary leaves the company, data both stored in Company X and in third parties must be deleted. The deletion of data from third parties is formalised as:

$$\{mO\delta \wedge hR\Phi\delta \wedge bXc\delta\}$$
$$\mathbf{delete}^{\dagger}(m, h, \delta)$$
$$\{\forall c.\langle hXc\delta \wedge cL\delta\rangle\ \mathbf{delete}^{\dagger}(b, c, \delta)\ \langle \neg cL\delta \wedge \neg cP\delta \wedge \neg cS\delta \wedge \neg cR\Phi\delta \wedge \neg bXc\delta\rangle\}$$

The action that deletes Mary's data from Company X is formalised as:

$$\{mO\delta \wedge hR\Phi\delta \wedge hL\delta\}$$
$$\textbf{delete}(m, h, \delta)$$
$$\{\neg hL\delta \wedge \neg hP\delta \wedge \neg hS\delta \wedge \neg hR\Phi\delta\}$$

Notice that the sequence which the actions will take place is irrelevant, since the result always remains the same.

5.3 Synopsis of the chapter

Every attempt to develop a novel language requires a verification process to evaluate the final result. The methodology which was deemed appropriate to reach the desired level of verification was the case study approach. According to the literature presented in Chapter 3, data subjects' privacy expectations are crafted based on power asymmetries which they experience when interacting with data controllers. There exist three different types of power asymmetries. In order to capture all the possible expectations of consent and revocation for systems implemented in different contexts, thus serving different business models with different legacy requirements in place, I applied the logic on three different case studies.

In this Chapter, the first attempt to verify the expressiveness of the logic is illustrated. The case study to provide the specification requirements for the consent and revocation controls was the Employee case study, which was the pilot study where the first primitive form of logic was tested. The aim of the chapter, which was successfully achieved, served a dual purpose. The first aim was to demonstrate that the ambiguities, emerging from the interpretation of the natural logic to machine-readable format and from the complex nature of privacy when the first version was applied, were tackled. The second was to verify that all the effects of consent and revocation preferences in a system offering controls were adequately formalised without any further reconfiguration of the logic.

More specifically, through a number of different use cases, different expectations of consent and revocation were effectively formalised, giving the opportunity to demonstrate examples of various Φ variables and obligations. In addition, with the design of the fictional WorkBook, the power asymmetries of the first kind, between two data subjects were also examined. To conclude, the final version of the logic presented in Chapter 4 was successfully applied to the Employee case study and

all the requirements were unambiguously formalised, without any further issues emerging.

CHAPTER 6

The Biobank case study

Following the formalisation of the pilot case study which verified the effectiveness of the logic in an environment where data subjects disclose personal data to various third parties and share data amongst them, the second case study where the logic is applied on is a Biobank based in the United Kingdom. The purpose of this case study is to formalise the diverse consent and revocation preferences that occur in a context totally different from the pilot case study, in order to test the expressiveness of the logic. The focus is on medical data treated in the Biobank and on the power relations between the patients and the researchers that are in favour of the researchers. Patients have the sense that their hopes for receiving a treatment depends on their participation. More issues emerge from the existence of the legacy system in place, whose requirements need to be cater for in the formalisations.

In the following Section 6.1, I present the details of the Biobank case study and the methodology followed to elicit the requirements for consent and revocation controls in this environment. Section 6.2, describes the use case in italics and illustrates their formalisations followed by explanatory text for each of them. The final Section 6.3, concludes the findings of this Chapter.

6.1 Requirements elicitation

A Biobank is a "resource of tissue and blood samples donated by patients for use in medical research" [199]. As a result, Biobanks collect and store samples in accordance with regulatory requirements and provide access to researchers in order to improve diagnosis and treatment, and ultimately patient care. The Biobank case study offers interesting issues in terms of managing consent and revocation controls in a context where sensitive information is handled, whereas legislation

imposes strict controls and patients' preferences need to be addressed. The aim is to verify that the logic is rich and expressive enough to allow the formalisation of requirements in a different context without adding new actions and rights.

I analyse a number of use cases that provide an overview of the environment where a system offering consent and revocation controls will be implemented. From these use cases, a list of requirements to explore the implications of invoking consent and revocation controls is elicited. The use cases that are formalised in this chapter are:

- The IT administrator creates consent and revocation options that will be presented to the patient both for the sample and the data.
- The IT administrator creates privacy access control policies.
- The IT administrator creates privacy obligation policies.
- The IT administrator sets consent and revocation default choices.
- The data subject (patient) or technician makes consent and revocation choices for specific study/studies.
- The technician is registering a sample and/or personal data in a spreadsheet.

The requirements regarding the consent and revocation functionality for the use cases in the Biobank case study were elicited from descriptions of the current legacy system in place and further information provided by the focus groups. The content analysis [154; 126; 190; 81] methodology was adopted to analyse the transcripts from the focus groups. The themes used in the content analysis were generated based on the consent and revocation actions defined in the logic which emerged from the conceptual consent and revocation model and from the application of the logic to the first case study. Both researchers associated with the biobank and patients participated in these focus groups.

6.2 Formalising the Biobank case study

This section illustrates the application of the logic to the use cases and provides evidence that the ambiguities identified and tackled in Chapter 5, do not emerge in the Biobank case study. In the formalisations below the letter a is used to denote

the patients of the Biobank, the letter b to denote the Biobank itself and the letter c to denote a researcher.

The data handled in this case study is hierarchised by creating three different domains in order to capture generic requirements (delete all data about my sample) by applying a single action to more that one datum, in a similar rationale to the one described in [43]. Each datum δ can only be allocated in one domain. The letter δ refers generally to all data that may exist in this case study. In addition, three different domains are created and all δ data that may occur will belong only to one of these domains. The letter δ_1 denotes the physical sample, the letter δ_2 the data derived from the sample and any measurements undertaken in the Biobank regarding the sample and the letter δ_3 denotes any personal data of the patient such as demographic data, name etc.

6.2.1 The IT administrator creates consent and revocation options that will be presented to the patient

In this use case the administrator, using the consent form designed by researchers and approved by the Research Ethics Committee (REC), writes the set of consent and revocation options to be offered to patients. The options derived from the current consent form that the patient needs to sign before the donation of the sample to the Biobank, and from the results of the focus groups. Only the actions available that could be triggered in the system are described.

Based on the analysis of the data generated by focus groups, the options available to patients concern the purpose for which the sample is given, the notification process, the ability to revoke permissions and delete data deriving from the sample and the request for the destruction of the sample. Thus, the patient may give consent to the Biobank to store, process and share data, constrain these choices by specifying the purpose of use and the parties that the data/sample will and will not be shared with. Furthermore, the patient could set notification preferences. It is not decided yet by the Biobank, whether there will be permitted an option enabling the patient to delegate consent to the next of kin and if the patient will be able to update some of her/his data.

Define the options for sharing data with researchers

The first option that a patient may express is to allow the Biobank to share the sample to researchers and provide restrictions regarding the purpose of the research, the background of the researchers and the duration of consent. This is formalised as:

$$\mathbf{grant}^*(a, b, \Phi, \delta)$$

where Φ = destination:$\Pi \wedge \neg\pi \wedge$ purpose:$u \wedge$ time duration:$t \wedge$ times processed:t^* and $\Pi \subseteq$ {Pharmaceutical companies, University}, $\pi \subseteq$ {Dr. Evil}, $u \subseteq$ {teaching, cancer research, DNA}, $t \in$ {One year - 40 years} and $t^* \in$ {One time - 100 times}.

Notice that I use the letter δ for data, thus the options refer to all the available data. More options could be provided to the patients to choose from and apply these options only for a specific domain of data e.g the purpose for which the sample will be used for.

With this formalisation the patient may choose to share data with researchers working in university laboratories, with researchers working for pharmaceutical enterprises but not with researchers working for Dr. Evil. Furthermore, they could control the purpose of the research and choose to share their data with researchers for teaching purposes, for cancer research, and to allow DNA analysis. The variable Π includes the parties that the Biobank is allowed to share data with, the variable π the parties that the Biobank is not allowed to share data with and the variable u describes the purposes for which data may be shared. The Biobank may also provide the option to the patient to choose the duration of consent and how many times the data/sample may be processed.

Define revocation and deletion options

Four different classes of revocation, symmetrical to what the patients give consent to, are distinguished. These options could affect the Biobank, the researcher or both and could be enabled either with a prospective or retrospective effect. It is worth noting that these options impact the business model of the Biobank and there might exist situations where the option of revocation would be prohibited due to legislation. The options for prospective revocation that a patient may ask for are presented below:

- Revoke permission to process sample /data after the Biobank has finished processing* it.

*Since Biobanks are repositories of samples, the term processing refers to the measurements

- Revoke permission to share sample/data from the Biobank after the Biobank has finished processing it.
- Revoke permission to process sample/data both from the Biobank and the researchers after the completion of the research.
- Destroy the sample/ Delete data.

Based on the same rationale, the different options of retrospective revocation offered are:

- Revoke permission to process sample /data from the Biobank before the Biobank has finished processing it.
- Revoke permission to share sample/data from the Biobank before the Biobank has finished processing it.
- Revoke permission to process sample/data both from the Biobank and the researchers before the completion of the research.
- Destroy the sample/ Delete data after the completion of the research.

The options of revoking consent and destroying/deleting the sample/data are formalised below. Whether the act of revoking consent will have a retrospective or prospective effect, is defined by the value of the variable p. If the variable is true the revocation is retrospective, otherwise the revocation is prospective.

With the option below, the patient revokes the right to process the data/sample from the Biobank. If the Biobank is in the process of transferring the sample/data to other researchers they should request it back. Samples and data shared to researchers previous to that action are not influenced. The option is formalised as:

$$\mathbf{revoke}(a, b, \Phi, \delta)$$

where $\Phi =$ currently processed : p and $p = \{true\}$.

With the option below, the patient revokes from Biobank the right to share data/sample. If the Biobank is in the process of transferring the sample/data to other researchers they should request it back. The subtle difference with the previous formalisation is that in this case the Biobank may still process the data/sample.

undertaken in the Biobank to specify the meta-data of the sample (Disease etc.)

Samples and data shared to researchers previous to that action are not influenced.The option is formalised as:

$$\mathbf{revoke^1}(a, b, \Phi, \delta)$$

where Φ = currently processed : p and $p = \{true\}$.

With the option below, the patient revokes from the researchers who collaborate with Biobank the right to process the data/sample. Samples and data shared to researchers prior to that option should not be processed further. The option is formalised as:

$$\mathbf{revoke}^{\dagger}(a, b, c, \Phi, \delta)$$

where Φ = currently processed : p and $p = \{true\}$.

With this option below the sample/data that is stored to Biobank is deleted/destroyed.

$$\mathbf{delete}(a, b, \Phi^*, \delta_2)$$

where Φ^* = disposal:$x \wedge$ currently processed : p and $x \subseteq$ {Destroy sample, delete data}and $p = \{true\}$.

The action that requires from the researchers to delete the acquired samples/data is:

$$\mathbf{delete}^{\dagger}(a, b, c, \Phi^*, \delta_2)$$

where Φ^* = disposal:$x \wedge$ currently processed : p and $x \subseteq$ {Destroy sample, delete data}and $p = \{true\}$.

A retrospective revocation is described, as the patient wishes to revoke rights while the data is currently processed by inserting the variable p. Furthermore, the variable x has been introduced to provide the patient with the opportunity to decide either to destroy the sample or to delete the data that derived from that sample.

Some of the revocation actions create obligations and all of them require actions to happen in the future before the patient may invoke them. For example, a patient cannot revoke permission from researchers to process their sample unless the Biobank has initially shared their sample with them.

Change of consent

The patient may also decide to change his initial consent. In this formalisation the change of restrictions is captured in the patient's initial consent. For example,

changing the time allowed to process data to 5 months:

$$\textbf{change}(a, b, \Phi, \delta,)$$

where Φ = use by:t and $t = 150$.

Set notification options

There is also the option for a patient to be notified under certain conditions.

$$\textbf{setnotify}(a, b, \Phi, \delta)$$

where Φ = notify-how:$n^* \wedge$ notify-what:$n^\dagger$ and $n^* \subseteq$ {email, general practitioner, Biobank website} and $n^\dagger \subseteq$ {results of the research, implication for health, new research, sample dispatched to researcher, sample destroyed}.

In this formalisation the patient may choose to be notified either by email, or via their GP or via the Biobank's website. Furthermore, the patient may choose to be notified when the results of the research are published, if the researchers by examining the sample believe that there could be further implications to her/his health, when the sample is dispatched for research or when it is destroyed.

IT administrator creates Biobank privacy access control policies

In order to create Biobank privacy access control policies the administrator needs to: define, or select one of the templates offered by the Biobank, and deploy them into the system. The privacy policies will be created from the allowed actions and the variables that will provide further constraints and information regarding the implementation of those actions. The suggested policies for the Biobank case study are:

1. I {consent/revoke consent} for Biobank to {collect/store/use} my personal data for { any research (provided it has been approved by Biobank and met all ethical standards of research); DNA specific research; selected clinical trials [list]; not at all} with access by {the research team that contacts me; pharmaceutical companies; others} (subject to time constraints/notification constraints).

2. I {consent/do not consent/revoke consent} for Biobank to {collect/store/use} my {sample and associated digital representations} for {Specified purpose}

(subject to time constraints/notification constraints).

3. I {consent/do not consent/revoke consent} for Biobank to share my sample (or its digital representations) for {Specified purpose} to {direct contacts of the researcher, anyone}.

4. I {consent/do not consent/revoke consent} for Biobank to share data for {any research (provided it has been approved by Biobank and met all ethical standards of research); selected clinical trials [list]; only the research team that contacts me}.

In the logic all the actions and the variables create a policy. The policy describes how the system will cope with each action and each variable and the patients' choices will define the values of the variables.Thus, the option of

$$\mathbf{grant}^*(a, b, \Phi, \delta)$$

where Φ = destination:$\Pi \wedge \neg\pi \wedge$ purpose:$u \wedge$ time duration:$t \wedge$ times processed:t^* and $\Pi \subseteq$ {Pharmaceutical, University}, $\pi \subseteq$ {Dr. Evil}, $u \subseteq$ {teaching, cancer research,DNA}, $t \in$ {One year - 40 years} and $t^* \in$ {One time - 100 times} is converted into a policy as: I {consent} for Biobank to {share} my sample for {teaching, cancer research, DNA} to {Pharmaceutical, University} and not to {Dr Evil} for the next {100 years} or being processed {100} times.

Each of the actions described in the previous section will create a separate policy with the same rationale. Some of the options include obligations which will be formalised in the section below.

The IT administrator creates Biobank privacy obligation policies

To create Biobank privacy obligation policies the administrator defines a set of obligation policy templates (and/or uses the templates offered by the Biobank).

The privacy obligation policies are perceived in two different ways in the logic:

- Those created to ensure that upon completion of an action, another action will be triggered in the future (notification requirements, delete data after 5 years).

- Those created by the obligation of a data controller to request further action from third parties with whom data has been shared, for the initial action to be

completed (request to propagate consent and revocation changes to all parties that process data).

The form of obligation policies for notification requirements is:

1. I {consent/do not consent/revoke consent} for Biobank to contact me about my data or sample via {email, phone, post, GP} when {my sample is shared, results of the research have gone public}.

The option for notification formalised as:

$$\mathbf{setnotify}(a, b, \delta, \Phi)$$

where Φ = notify-how:n^* $\wedge$ notify-what:$n^\dagger$ and $n^* \subseteq$ {email, general practitioner, Biobank website} and $n^\dagger \subseteq$ {results of the research, implication for health, new research, sample dispatched to researcher, sample destroyed} will be converted into an obligation policy as: I {consent} for Biobank to contact me about my data or sample via {email, general practitioner, Biobank website} when {results of the research are finalised, implication for health, new research, sample dispatched to researcher, sample destroyed}.

The IT administrator sets consent and revocation default choices

To set the default consent and revocation choices the administrator accesses the "admin tool" box which propagates configuration changes to required components.

$$\{aO\delta_2 \wedge bR^\dagger\Phi\delta_2\}$$
$$\mathbf{grant^1}(a, b, \Phi, \delta_2)$$
$$\{bL\delta_2 \wedge bP\delta_2 \wedge bS\delta_2 \wedge bR\Phi\delta_2\}$$

where Φ = destination:Π $\wedge$ $\neg\pi$ $\wedge$ purpose:u $\wedge$ time duration:t $\wedge$ times processed:t^* and $\Pi \subseteq$ {Pharmaceutical, University}, $\pi \subseteq$ {Dr. Evil}, $u \subseteq$ {teaching, cancer research, DNA}, $t \in$ {One year - 40 years}, $t^* \in$ {One time - 100 times} and δ_3 is

data concerning the patient's profile (registration number, further information).

6.2.2 The data subject (patient) or Biobank technician makes consent and revocation choices for specific study/studies)

For specific studies the data concern the sample only. A new variable could also be introduced to denote that the patient is participating in a study and he/she requires more options regarding, for example, the purpose of using the sample.

$$\{aO\delta_1 \wedge bR^{\dagger}\Phi\delta_1\}$$
$$\mathbf{grant^1}(a, b, \Phi, \delta_1)$$
$$\{bL\delta_1 \wedge bP\delta_1 \wedge bS\delta_1 \wedge bR\Phi\delta_1\}$$

where Φ = destination:$\Pi \wedge \neg\pi \wedge$ purpose:$u \wedge$ time duration:$t \wedge$ times processed:t^* and $\Pi \subseteq$ {Team that consults me}, $\pi \subseteq$ {Anyone else}, $u \subseteq$ {DNA}, $t \in$ {One year - 40 years}, $t^* \in$ {One time} and δ_1 is a sample collected due to the patients' participation in a specific study.

Patient registration

When a new patient (data subject) registers, or more likely an authorised Biobank employee is acting on his behalf by interacting with the Biobank's system, it will assign a new patient ID, and then a new ID all linked to the new trial. This action is formalised as:

$$\{aO\delta_3 \wedge bR^{\dagger}\Phi\delta_3\}$$
$$\mathbf{grant^1}(a, b, \Phi, \delta_3,)$$
$$\{bL\delta_3 \wedge bP\delta_3 \wedge bS\delta_3 \wedge bR\Phi\delta_3\}$$

where Φ = destination:$\Pi \wedge \neg\pi \wedge$ purpose:$u \wedge$ time duration:$t \wedge$ times processed:t^* and $\Pi \subseteq$ {Pharmaceutical, University}, $\pi \subseteq$ {Dr. Evil}, $u \subseteq$ {teaching, cancer research}, $t \in$ {One year - 40 years}, $t^* \in$ {One time - 100 times} and δ_3 is data regarding the patient's profile (registration number, further information).

When the patient registers with the Biobank she/he discloses personal data and

obtains a registration number. Upon this data, the patient has the options to restrict the usage of the data, the destinations to where the data will be shared and specify the time that data will be stored for.

During tissue sample collection and data entry

This use case is very similar to the one described in the section above, with the difference that before locally storing the data subject's sample, the Biobank technician is asked to define preferences for the new sample collected. When a Biobank technician collects a sample, he logs onto the system, selects the Samples tab and clicks Add Sample. He then fills in the required Sample details on the Sample Data Entry Page.

$$\{aO\delta_1 \wedge bR^{\dagger}\Phi\delta_1\}$$
$$\mathbf{grant^1}(a, b, \Phi, \delta_1)$$
$$\{bL\delta_1 \wedge bP\delta_1 \wedge bS\delta_1 \wedge bR\Phi\delta_1\}$$

where Φ = destination:$\Pi \wedge \neg\pi \wedge$purpose:$u \wedge$time duration:$t \wedge$times processed:$t^*$ and $\Pi \subseteq$ {Pharmaceutical, University}, $\pi \subseteq$ {Dr. Evil}, $u \subseteq$ {teaching, cancer research, DNA}, $t \in$ {One year - 40 years} and $t^* \in$ {One time - 100 times}.

Because it is a sample for trial the introduction of new variables may better describe the purpose of the research, such as DNA purposes. The patient may also choose to set notification requests.

6.2.3 The data subject (patient) or Biobank technician changes consent and revocation choices for specific study/studies

In this use case the Biobank technician logs into the system and changes on the patient's behalf her/his choices.

$$\{aO\delta_1 \wedge bR^{\dagger}\Phi'\delta_1 \wedge bR\Phi\delta_1\}$$
$$\mathbf{change}(a, b, \Phi', \delta_1)$$
$$\{\neg bR\Phi\delta_1 \wedge bR\Phi'\delta_1\}$$

where Φ' = destination:$\Pi \wedge \neg\pi \wedge$ purpose:$u \wedge$ time duration:$t \wedge$ times processed:t^* and $\Pi \subseteq$ {University} and $u \subseteq$ {DNA}, $t \in$ {One year} and $t^* \in$ {One time} $\implies$ obligation.

With the action above the patient chooses to change her/his consent regarding the sample and to limit the parties having access to it, the duration of consent and the times that it is allowed to be processed. This action implies an obligation because when the existing consent is no longer valid the patient should choose to delete his/her sample or to change his consent.

6.2.4 Tissue sample collection, data and consent and revocation choices entry

When the sample is collected by the Biobank, the patient expresses specific consent and revocation choices that should be enforced and respected by the Biobank and any researcher that may acquire the sample.

$$\{aO\delta \wedge bR^{\dagger}\Phi\delta\}$$
$$\mathbf{grant^1}(a, b, \Phi, \delta)$$
$$\{bL\delta \wedge bP\delta \wedge bS\delta \wedge bR\Phi\delta\}$$

where Φ = destination : $\Pi \wedge$ purpose:u and Π = {University} and $u \subseteq$ {cancer research, DNA} $\implies$ obligation.

In the above formalisation the patient donates a sample to the Biobank. The pre-condition declares that the donor of the sample is also the owner of the sample ($aO\delta$), and that the Biobank must be willing to accept their consent and revocation choices ($bR^{\dagger}\Phi\delta$). The sample is then registered in the Biobank. As a result in the post condition the Biobank has stored the sample ($bL\delta$) and it may process it ($bP\delta$), share it ($bS\delta$) but must always respect the restrictions the patient has imposed. In this case, the sample may only be shared with the university laboratories, specifically for cancer research purposes and DNA analysis.

6.2.5 The Biobank technician is sharing a sample and/or personal data in a spreadsheet

In this use case, the Biobank is sharing data about a sample with a researcher in digital form. A researcher requests measurements of data from a sample and the Biobank provides the data in a spreadsheet. This is personal data, so all the controls which a patient may have imposed on the sample, should be passed on.

$$\{bS\delta \wedge bR\Phi\delta \wedge c \in \cup\Phi.\text{destination} \wedge \Phi^* \leq \Phi \wedge cR^\dagger\Phi^*\delta\}$$
$$\mathbf{grant}(b, c, \delta, \Phi^*)$$
$$\{cL\delta \wedge cP\delta \wedge cR\Phi^*\delta \wedge$$
$$\langle bNa\delta\rangle\ \mathbf{setnotify}(a, b, \Phi', \delta)\ \langle true\rangle$$
$$\wedge\forall c.\langle bNa\delta \wedge aN^\dagger\delta \wedge \text{“}shared\text{”} \in \cup\Phi.\text{notify-what}$$
$$\wedge\text{“}email\text{”} \in \cup\Phi.\text{notify-how}\rangle\ \mathbf{notify}(b, a, \delta, \text{“}shared\text{”}, \text{“}email\text{”})\ \langle true\rangle\}$$

where $\Phi^* = \text{purpose} : u$ and $u = \{\text{cancer research}\}$.

Note that the use of an action which only allows the researcher to process the data but not to share it is preferred. Also, all the controls which are appropriate for the right to process, are cascaded to the researcher. Furthermore, as the patient had set notification requirements, an email is sent to notify the patient that the data has been shared with the researcher.

The focus groups highlighted the importance of notification in such cases. If notification choices are available, then there is an obligation triggered requiring from the data controller to notify the patient. It is crucial to make a decision which will determine the Biobank's strategy regarding the notification process. Possible options could be notification by email, using a link on the Biobank's website pointing to the published papers of the researchers, or requesting the contribution of the patient's GP.

There will be cases where the consent given by the patient will not be specific enough for the Biobank to determine whether the sample should be shared or not. Therefore, the consent will be implied. The patient could decide whether she/he wishes to be informed if such circumstances occur. Thus, depending on her/his choices she/he might choose to allow sharing of data by default or to be asked for approval. The focus groups conducted with researchers pointed out that any notification may disturb the patients and their family, while on the other hand enabling

researchers to continue their research by using the same samples for different purposes is of significant importance. There are conflicting needs to be catered for and to-date the Biobank has not resolved the issue. Unfortunately, there does not exist any functionality to handle the notification process. The standard procedure requires researchers to get in conduct with the patients. However, it is a time-consuming process and most of the times the patients never reply to researchers' emails. Instead of annoying patients by allowing researchers to conduct them, the Biobank could incorporate the functionality described in the use cases, to allow them to explicitly specify under which circumstances the Biobank may get in conduct with them.

6.3 Synopsis of the chapter

Following the successful formalisation of the Employee case study with the application of the consent and revocation logic, this chapter sought to raise new challenges and novel consent and revocation expectations. A context different from the first case study, since it required the elicitation of different requirements not only for the data subjects but for the legacy system which handled personal data as well, was chosen and the Biobank was considered the ideal case study to question the applicability of the logic. The power asymmetries regarding the Biobank are different from the first case study, since the Biobank's patients believe that they reside their hopes for a treatment on their participation to new trials. Thus, they feel that they are in a disadvantageous position when it comes to give consent to disclose data, a factor that is depicted in their consent revocation expectations and enhanced the decision to choose the specific case study.

The aim of the chapter was to validate the expressiveness of the logic in a different case study, in a manner that would not raise further ambiguities or require any further refinements. Based on the focus groups conducted by the EnCoRe project that were specifically designed for the Biobank case study, I gained a better understanding of the environment which the controls should be invoked. The generated data from the focus groups was analysed adopting the same methodology followed for the development of the conceptual model and facilitated the process of eliciting requirements. Thus, resulting in more detailed consent and revocation options available to the patients with respect to the requirements elicited from the analysis of the legacy system in place and the demands of the recipients of the medical data

who are the researchers of the Biobank.

The application of the logic to the demanding second case study, enhanced my confidence regarding the richness of the logic and it was the first successful step to demonstrate the capability of the logic to formalise requirements in different contexts, for different business models and for different data subjects' expectations. The formalisation was concluded without any ambiguities emerging and all the specifications for the legacy system were effectively captured. It is worth noting that new variables were used to capture the changes in the context, providing evidence that allowing the defining of new variables increases the flexibility and the expressiveness of the logic, while retaining its rigid and correct mathematical form.

CHAPTER 7

The Identity Assurance Programme case study

Having established the richness of the logic in two different contexts, where no further ambiguities emerged and no need for a further refinement of the logic presented in Chapter 4 occurred, the next challenge is to apply the logic in an environment where data subjects are requesting to express controls on data disclosed to governmental services. The Identity Assurance Programme case study, serves the purpose for that task. As demonstrated in Chapter 3, data subjects respond to the power asymmetries that occur when they interact with the governmental sector, by perceiving privacy preferences differently. In addition, the business model followed in the case study differs from the business models of the aforementioned formalised case studies, raising new challenges for consent and revocation controls. The aim of this Chapter is to provide evidence that the logic is expressive enough to adjust to the power asymmetries and formalise the specification of requirements adequately, without any further ambiguities emerging.

The next Section 7.1 provides the description of the Identity Assurance Programme case study and explains the methodology followed to elicit the requirements for this environment. In Section 7.2, the descriptions of the use cases are presented in italics, followed by the formalisations and their explanation. The final Section 7.3 provides a summary of the results presented in this Chapter.

7.1 Requirements elicitation

The Identity Assurance Programme case study describes an identity service concept facilitating a government to certify the identity of data sub-

> jects. It is based on the development of a consistent, customer-centric approach to digital identity assurance across all public services. This will allow service users to log on safely to digital public services in a way that ensures personal privacy, reduces fraud and facilitates the move to online public services.
>
> Online services have the potential to make life more convenient for service users as well as delivering cost savings. However, currently customers have to enter multiple log-in details and passwords to access different public services, sometimes on the same website. This involves significant duplication, is expensive to operate and is highly inconvenient for users. It acts as a deterrent to people switching to digital channels, hampers the vision of digital being the primary channel for accessing Government information and transactions, and provides an opportunity for fraudsters.
>
> The intention is to create a market of accredited identity assurance services delivered by a range of private sector and mutualised suppliers. A key improvement will be that people will be able to use the service of their choice to prove identity when accessing any public service. Identity assurance services will focus on the key imperative to ensure privacy [252].

To mitigate the risks associated with transacting online, almost all services require the user to go through some form of initial registration and subsequent login procedure. These procedures need to acquire consent from the citizens.

As summarised in the EnCoRe internal report "3^{rd} Case Study user's Requirements Workshops" [184], the principles of the Identity Assurance Programme include the following statements:

- Customer focus: an identity assurance solution must be based around the needs of the individual otherwise it will not be used or valued. People have different and changing needs. No one "big brother" solution will meet the needs of all customers in all contexts.
- Customer control: the use of a customer's identity and personal data should be fully transparent and controlled by the customer.

These are core principles that must be fully supported by a system when the goal is to enable data subjects (customers, citizens) to express and change their consent

(on their personal data) and organisations to explicitly enforce the management of consent and privacy preferences.

Figure 7.1 below illustrates how a system, and in this case EnCoRe, can be deployed within a business model addressing identity assurance issues and how it influences the key players in the environment (Identity Providers, Attribute Providers, Hub, Service Providers) providing the required consent and privacy management functionality [184]:

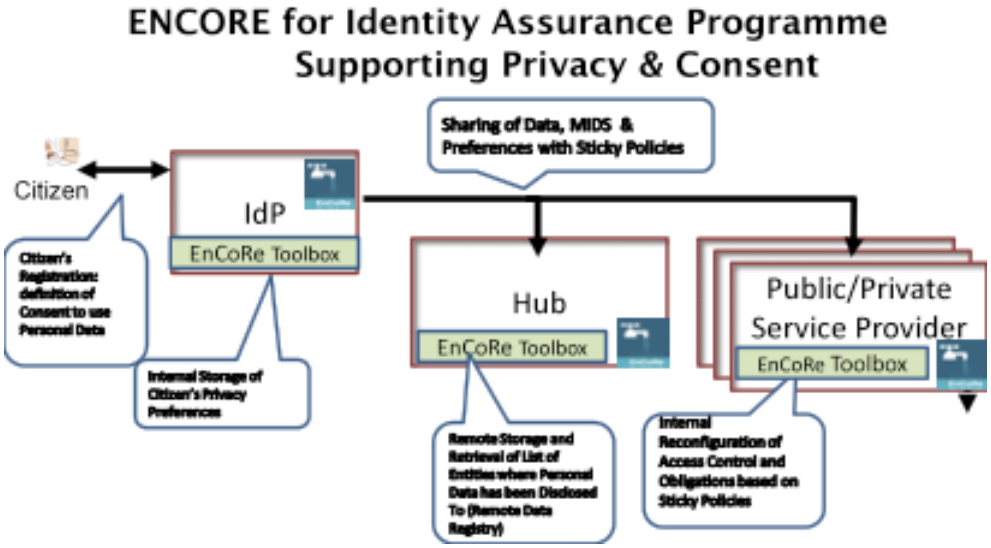


Figure 7.1: EnCoRe in the Identity Assurance Programme architecture. A picture created by HP laboratories in Bristol [184]

The use cases described in this case study aim to address the problem of handling dynamic consent and therefore enable the benefits of public services delivered through digital channels to reach all parts of society, with respect to citizens' privacy. All the key players of the system are formalised in the use cases. The aim is to verify that the logic can be applied to formalise all the requirements without any additional actions and without any further ambiguities emerging.

Eliciting requirements for the Identity Assurance Programme raises several challenges. The business model is still a matter of debate and the project's proposition was not definitive at the time when I wrote the thesis. Thus, the requirements were elicited from analysing documents and discussions from the research community which was assigned to assist those people that initiated the Identity Assurance Programme.

7.2 Formalising the Identity Assurance Programme case study

A number of use cases is identified that provide an overview of the environment where the system will be implemented and a list of requirements to explore the implications of invoking consent and revocation controls is elicited from these use cases. The use cases which are formalised in this section are driven from discussions among the research community which is participating in the Identity Assurance Programme.

7.2.1 Semantics of the use cases

There are three different types of data subjects that could request access to public services:

- a citizen or individual customer
- an employee of a business
- an agent authorised to act on behalf of a person or business

The symbols used in the formalisation of the following use cases are:

- a denotes a data subject (an individual, an employee of a business or an agent authorised to act on behalf of a person or business).
- ip denotes an Identity Provider.
- ap denotes an Attribute Provider.
- sp denotes a Service Provider.
- h denotes the Hub
- δ represents a Minimum Identity Data Set (MIDS). MIDS, for the needs of this case study, consists of:
  - Identifier
  - First name
  - Surname
  - Date of birth
  - Gender

  - House number
  - Post Code

- δ_1 represents further attributes provided by the Attribute Providers.
- k is next of kin
- c is a business which requests the services of the Service Providers

7.2.2 Use cases related to consent and revocation

Customer registers with an Identity Provider

An Identity Provider is requested to create a "digital identity" by a data subject, that will be used to acquire access to public services. The identity is produced, since the Identity Provider verifies the credentials of the citizen to the acquired standards.

$$\{aO\delta \wedge ipR^{\dagger}\Phi\delta\}$$
$$\mathbf{grant}^{*}(a, ip, \Phi, \delta)$$
$$\{ipL\delta \wedge ipP\delta \wedge ipS\delta \wedge ipR\Phi\delta\}$$

where Φ = destination : $\Pi \wedge$ purpose:$u \wedge$ time:t and $\Pi = \{\text{Hub}\}$, $u = \{$to be shared with the hub$\}$, t= 1 year $\implies$ obligation.

In this formalisation the data subject has given consent to the Identity Provider to process and share the "MIDS" only with the hub and for the period of one year. The time restriction implies an obligation for the system, which will be triggered after the designated time has elapsed. A variable that will allow the data subject to define which action should be triggered by the obligation (prompt to re-consent, delete data etc) can be included.

Pre-condition variations:

- the customer:
  - is under 18 years of age. In this case a variable g is added to denote age. When the age of the data subject is less than 18 years old an obligation is created to request consent of the next of kin. This is formalised as follows:

$\langle idL\delta \wedge g \leq 18 \rangle$ **delegate**(a, k, Φ, δ) $\langle kU\delta \rangle$

– is non-British. It is similar to the first formalisation. An extra variable *nationality* is included to denote the nationality of the data subject. Thus the formalisation is:

$$\{hS^*\delta \wedge ipR^\dagger\Phi\delta\}$$
$$\textbf{grant}^*(h, ip, \Phi, \delta)$$
$$\{ipL\delta \wedge ipP\delta \wedge ipS\delta \wedge ipS^*\delta \wedge ipR\Phi\delta\}$$

where Φ = destination : $\Pi \wedge$ purpose:$u \wedge$ time:$t \wedge$ nationality : *nationality* and Π = {Hub}, u = {to be shared with the Hub}, t = 1 year and *nationality* = { Greek } $\implies$ obligation.

Customer uses an Identity Provider to access a public service

A customer wishes to access a public service and is first asked to provide a "digital identity" from an accredited Identity Provider. The customer authenticates to the Identity Provider who then provides the customer's minimum identity data set to the public service provider. The customer is then able to access the public service. As the customer has not used the Identity Provider to access the public service on a previous occasion, the public service requires some additional piece of information before it can provide access to the service. Once this has been provided then the customer is then able to access the public service.

The assumption in this use case is:

- The data provided by the Identity Provider is not sufficient to access the public service automatically.

The formalisation of this use case is as follows:

$$\{aO\delta \wedge spR^{\dagger}\Phi'\delta\}$$
$$\textbf{request}(sp, h, \Phi', \delta)$$
$$\{spR\Phi'\delta\wedge$$
$$\langle hR^{\dagger}\Phi'\delta\rangle\ \textbf{request}(h, ip, \Phi', \delta)\ \langle hR\Phi'\delta\wedge$$
$$\langle hR^{\dagger}\Phi \wedge ipS^{*}\delta\rangle\ \textbf{grant}^{*}(ip, h, \Phi, \delta)\ \langle hL\delta \wedge hP\delta \wedge hS^{*}\delta \wedge hR\Phi\delta\wedge$$
$$\langle hS^{*}\delta \wedge spR^{\dagger}\Phi\delta\rangle\ \textbf{grant}(h, sp, \Phi, \delta)\ \langle spR\Phi\delta \wedge spL\delta \wedge spP\delta\rangle\rangle\rangle\}$$

where Φ = destination : $\Pi \wedge$ purpose:$u \wedge$ time:t and $\Pi = \{\text{Hub}\}$, $u = \{$to be shared with the Hub$\}$, $t = 1$ year $\implies$ obligation.

Φ' = purpose:u and $u = \{$for electoral purposes$\}$.

In this use case, the user is requested to provide evidence of their identity in order to access a public service. Through the service provider's website she/he chooses an Identity Provider that has already verified their identity. The hub is the mediator who forwards the request of the user to the Identity provider and then the data from the Identity provider to the service provider. Notice that the Service provider is obliged to conform with both Φ and Φ' restrictions and that the hub has the ability to understand and forward the policies of the Identity Providers or the Attribute Providers and not the ability to store or "understand" the data subjects' data.

Customer uses an Identity Provider to access a public service but must provide additional information

The public service requires some additional piece of information before it can provide access to the service. Once this has been provided then the customer is able to access the public service. First the customer is asked to provide a "digital identity" from an accredited Identity Provider. The customer authenticates to the Identity Provider who then provides the customer's minimum identity data set to the public service provider.

The formalisation of this use case is:

$$\{aO\delta \wedge spR^{\dagger}\Phi'\delta\}$$
$$\mathbf{request}(sp, h, \Phi', \delta)$$
$$\{spR\Phi'\delta\wedge$$
$$\langle hR^{\dagger}\Phi'\delta\rangle\ \mathbf{request}(h, ip, \Phi', \delta)\ \langle hR\Phi'\delta\wedge$$
$$\langle hR^{\dagger}\Phi \wedge ipS^{*}\delta\rangle\ \mathbf{grant}^{*}(ip, h, \Phi, \delta)\ \langle hL\delta \wedge hP\delta \wedge hS^{*}\delta \wedge hR\Phi\delta\wedge$$
$$\langle hS^{*}\delta \wedge spR^{\dagger}\Phi\delta\rangle\ \mathbf{grant}(h, sp, \Phi, \delta)\ \langle spR\Phi\delta \wedge spL\delta \wedge spP\delta\rangle\rangle\rangle\}$$

where Φ = destination : $\Pi \wedge$ purpose:$u \wedge$ time:t and $\Pi = \{\text{Hub}\}$, $u = \{$to be shared with the Hub$\}$, t = 1 year $\implies$ obligation and Φ' = purpose:u and $u = \{$for electoral purposes$\}$.

But since "MIDS" is not enough to allow the data subject to acquire access to the services, there is a request send to the attribute provider to disclose further information.

$$\{aO\delta_1 \wedge spR^{\dagger}\Phi'\delta_1\}$$
$$\mathbf{request}(sp, h, \Phi', \delta_1)$$
$$\{spR\Phi'\delta_1\wedge$$
$$\langle hR^{\dagger}\Phi'\delta_1\rangle\ \mathbf{request}(h, ap, \Phi', \delta_1)\ \langle hR\Phi'\delta_1\wedge$$
$$\langle hR^{\dagger}\Phi \wedge apS^{*}\delta_1\rangle\ \mathbf{grant}^{*}(ap, h, \delta_1, \Phi'')\ \langle hL\delta_1 \wedge hP\delta_1 \wedge hS^{*}\delta_1 \wedge hR\Phi'\delta_1'\wedge$$
$$\langle hS^{*}\delta_1 \wedge spR^{\dagger}\Phi''\delta_1'\rangle\ \mathbf{grant}(h, sp, \Phi'', \delta_1)\ \langle spR\Phi''\delta_1 \wedge spL\delta_1 \wedge spP\delta_1\rangle\rangle\rangle\}$$

where Φ'' = destination : $\Pi \wedge$ purpose:$u \wedge$ time:t and $\Pi = \{\text{Hub}\}$, $u = \{$to be shared with the Hub$\}$, t= 2 years $\implies$ obligation.

Φ' = purpose:u and $u = \{$for electoral purposes$\}$.

In order for the action to be fulfilled, the attribute provider must have obtained consent for the verified data from the data subject. This is formalised as:

$$\{aO\delta_1 \wedge apR^{\dagger}\Phi'\delta_1'\}$$
$$\mathbf{grant}^{*}(a, ap, \delta_1, \Phi'')$$
$$\{apR\delta_1 \wedge apL\delta_1 \wedge apP\delta_1 \wedge apS\delta_1 \wedge apS^{*}\delta_1\}$$

where Φ'' = destination : $\Pi \wedge$ purpose:$u \wedge$ time:t and $\Pi = \{\text{Hub}\}$, $u = \{$to be shared with the Hub$\}$, t= 2 years $\implies$ obligation.

Notice that in this use case the Service Provider should accept data subject's choices of consent Φ that dictate the data received from the Identity Provider and those Φ'' conditions that dictate the data received from the Attribute Provider. In addition, there are further restrictions Φ' which must be imposed to the overall data.

The customer wishes to move from another public service

Successfully having authenticated with an accredited Identity Provider in order to access a public service the customer now wishes to move to a separate public service provided by another Service Provider.

The formalisation of this use case is as follows:

$$\{aO\delta \wedge spR^{\dagger}\Phi'\delta\}$$
$$\mathbf{request}(sp_1, sp, \Phi', \delta)$$
$$\{sp_1R\Phi'\delta\wedge$$
$$\langle spS^*\delta\rangle\ \mathbf{grant}^*(sp, sp_1, \Phi, \delta)\ \langle sp_1R\Phi\delta \wedge sp_1P\delta \wedge sp_1S^*\delta\rangle\}$$

where Φ = destination : $\Pi \wedge$ purpose:$u \wedge$ time:t and $\Pi = \{\text{Hub}\}$, $u = \{$to be shared with the Hub and other service providers$\}$, t = 2 years $\implies$ obligation. and Φ' = purpose:u and $u = \{$for taxation$\}$.

Notice that the user must have consented to the service provider to share the data with another service provider. If the hub mediates to complete the transaction then the request is made to the hub and the formalisation is similar to the one described in the 7.2.2 use case.

Customer creates, modifies or removes an authority for another user

A customer wishes to create, modify or remove authority for another person to conduct particular transactions with a public service.

In this use case the data subject decides to revoke the delegation rights. This is formalised as:

$$\{aO\delta \wedge cU\delta\}$$
$$\textbf{revoke\_delegate}(a, c, \delta)$$
$$\{\neg cU\delta\}$$

After the action the c company will not have access to the system to handle the consent procedures on behalf of the data subject.

A customer requires a proxy to access a public service on their behalf

The customer informs the public service provider of the contact details of their chosen proxy plus a shared secret that will be required by the proxy to complete the process.

This is formalised as:

$$\{aO\delta \wedge spR^{\dagger}\Phi\delta\}$$
$$\textbf{delegate}(a, c, \Phi, \delta)$$
$$\{cU\delta \wedge spR\Phi\delta\}$$

where Φ = willing to accept : w and $w = \{\text{c company}\}$. With this formalisation, c company may acquire access and act on behalf of the data subject. There is a restriction in the pre-condition that the Service Provider is willing to accept c as a delegated party.

Customer removes an authority for another user

A representative of an organisation modifies / suspends / revokes the authority of a representative of the organisation to act on behalf of the organisation in particular capacities.

$$\{aO\delta \wedge kU\delta\}$$
$$\textbf{revoke\_delegate}(a, k, \delta)$$
$$\{\neg kU\delta\}$$

Customer closes a relationship with an Identity Provider

A customer requests to close a relationship with an accredited Identity Provider and to ensure that all personal data is deleted as per agreed terms and conditions. Prior to closing the relationship, the Identity Provider will ensure that the customer is the real owner of the identity via credential authentication and proof of ownership of their contact details. The customer's appropriate transaction history is retained for audit purposes.

The formalisation for this use case is:

$$\{aO\delta \wedge ipL\delta \wedge ipP\delta \wedge ipS^*\delta \wedge ipR\Phi\delta\}$$
$$\mathbf{delete}(a, ip, \Phi', \delta)$$
$$\{\neg ipL\delta \wedge \neg ipP\delta \wedge \neg ipS^*\delta \wedge \neg ipR\Phi\delta\}$$

where Φ' = deletion : x and x= {retain data for audit purposes for as long as legislation permits}.

It is still undefined whether the Service Provider will request new data from another Identity Provider and what will happen in the event where further information has been disclosed by an Attribute Provider. There are two possible scenarios. In the first scenario the Service Provider is no longer offering the service to the data subject, rendering the information redundant. A possible solution could be to send a notification to the Attribute Service stating that the data is no longer processed and that it is deleted. This is formalised as:

$$\{aO\delta \wedge spL\delta \wedge spP\delta \wedge spS^*\delta \wedge spR\delta\Phi\}$$
$$\mathbf{delete}(a, sp, \Phi', /delta)$$
$$\{\neg spR\Phi\delta \wedge \neg spL\delta \wedge \neg spP\delta \wedge \neg spS^*\delta \wedge$$
$$\langle hXsp\delta\rangle\ \mathbf{notify}(sp, h, \delta)\ \langle \neg hXsp \wedge$$
$$\langle apXh\delta\rangle\ \mathbf{notify}(h, ap, \delta)\ \langle \neg apXh\rangle\rangle\}$$

where Φ' = deletion : x and x= {retain data for audit purposes for as long as legislation permits}.

In the second scenario, the Service Provider requests from the data subject to choose an alternative Identity Provider. In this case there is an obligation created

and the formalisation is similar to the 7.2.2 use case.

Customer closes a relationship with a Service Provider

A customer requests to close a relationship with a Service Provider and to ensure that all personal data is deleted as per agreed terms and conditions. Prior to closing the relationship, the Service Provider will ensure that the customer is the real owner of the identity via credential authentication and proof of ownership of their contact details. The customer's appropriate transaction history is retained for audit purposes.

The formalisation of this use case is:

$$\{aO\delta \wedge spL\delta \wedge spP\delta \wedge spS^*\delta\}$$
$$\mathbf{delete}(a, sp, \Phi', \delta)$$
$$\{\neg spL\delta \wedge \neg spP\delta \wedge \neg spS^*\delta \wedge$$
$$\langle hX sp\delta\rangle\ \mathbf{notify}(sp, h, \delta)\ \langle \neg hX sp \wedge$$
$$\langle apX h\delta_1\rangle\ \mathbf{notify}(h, ap, \delta_1)\ \langle \neg apX h\rangle$$
$$\wedge \langle ipX h\delta\rangle\ \mathbf{notify}(h, ip, \delta)\ \langle \neg ipX h\rangle\rangle\}$$

where Φ' = deletion : x and x= {retain data for audit purposes for as long as legislation permits}.

Before the termination of the service, a notification is sent by the hub to the Identity and Attribute Providers to notify that the data is not used by the Service Provider any more. It is crucial that the service provider is in a position to identify which data δ acquired from the *ip*s is related to which data $delta_1$ acquired by the *ap*s. The hub, although state-less must be able to bundle the preferences and the data from the *ip* and the *ap*.

7.2.3 Use cases loosely related to consent and revocation

Customer cannot establish a trustworthy identity with an Identity Provider

A customer requests an accredited Identity Provider to establish a "digital identity" to enable the customer to access a public service. The Identity Provider is unable to verify the customer's identity to the level of assurance required by the Service Provider.

This use case is not related to consent or revocation options but it is used to demonstrate the notification possibilities. The data subject may have chosen to be notified when a connection is not established.

$$\{aO\delta \wedge aN^{\dagger}\delta \wedge ipP\delta \wedge ipR\Phi\delta \wedge ipR^{\dagger}\Phi^{*}\delta\}$$
$$\textbf{setnotify}(m, ip, \Phi^{*}, \delta)$$
$$\{ipR\Phi^{*}\delta \wedge ipLn^{\dagger} \wedge ipLn^{*} \wedge ipNa\delta\}$$

where Φ^{*} = notify-how:$n^{*} \wedge$ notify-what:$n^{\dagger}$ and $n^{*} \subseteq$ {email, via hub} and $n^{\dagger} \subseteq$ {Unable to verify identity with the current data}.

Customer is already registered with the Identity Provider

A customer requests an accredited Identity Provider to establish a "digital identity" to enable the customer to access a public service. However, the customer's identity has already been registered with the Identity Provider.

$$\{aO\delta \wedge aN^{\dagger}\delta \wedge ipP\delta \wedge ipR\Phi\delta \wedge ipR^{\dagger}\Phi^{*}\delta\}$$
$$\textbf{setnotify}(m, ip, \Phi^{*}, \delta)$$
$$\{ipLn^{\dagger} \wedge ipLn^{*} \wedge ipR\Phi^{*}\delta \wedge ipNa\delta\}$$

where Φ^{*} = notify-how:$n^{*} \wedge$notify-what:$n^{\dagger}$ and $n^{*} \subseteq$ {email, via hub} and $n^{\dagger} \subseteq$ {The identity is already registered}.

Customer cannot access a public service using the identity from an Identity Provider

A customer wishes to access a public service and is first asked to provide a "digital identity" from an accredited Identity Provider. The customer authenticates to the Identity Provider who then provides the customer's minimum identity data set to the public service provider. However, the public service is unable to provide access to the service. A customer wishes to access a public service for which the customer has already established an account using an identity issued by an acceptable Identity Provider.

The customer accesses the public service and is asked to present valid credential(s) in order to be authenticated to the Identity Provider. The public service does not allow access to the service either because of a failure at the public service, or because of a failure at the Identity Provider.

In the first case, the customer is presented with an explanation that the service was unable to authenticate the credential, and the customer is directed to contact the Identity Provider.

$$\{aO\delta \wedge aN^{\dagger}\delta \wedge ipP\delta \wedge ipR\Phi\delta \wedge ipR^{\dagger}\Phi^{*}\delta\}$$
$$\textbf{setnotify}(m, ip, \Phi^{*}, \delta)$$
$$\{ipLn^{\dagger} \wedge ipLn^{*} \wedge ipR\Phi^{*}\delta \wedge ipNa\delta\}$$

where Φ^{*} = notify-how:$n^{*} \wedge$ notify-what:$n^{\dagger}$ and $n^{*} \subseteq$ {email, via hub} and $n^{\dagger} \subseteq$ {Unable to authenticate credentials}.

In the second case, the customer is presented with an explanation that the service is experiencing problems and the customer is directed to contact the service administrator through other alternatives.

$$\{aO\delta \wedge aN^{\dagger}\delta \wedge ipP\delta \wedge ipR\Phi\delta \wedge ipR^{\dagger}\Phi^{*}\delta\}$$
$$\textbf{setnotify}(m, ip, \Phi^{*}, \delta)$$
$$\{ipLn^{\dagger} \wedge ipLn^{*} \wedge ipR\Phi^{*}\delta \wedge ipNa\delta\}$$

where Φ^{*} = notify-how:$n^{*} \wedge$notify-what:$n^{\dagger}$ and $n^{*} \subseteq$ {email, via hub} and $n^{\dagger} \subseteq$ {The service is experiencing problems, you will be redirected to an alternative service}.

Customer cannot authenticate using his or her credentials

A customer wishes to access a public service and is first asked to provide a "digital identity" from an accredited Identity Provider. A customer wishes to access a public service and is first asked to provide a "credential" from an accredited Identity Provider. The customer's attempts to authenticate fail.

$$\{aO\delta \wedge aN^{\dagger}\delta \wedge ipP\delta \wedge ipR\Phi\delta \wedge ipR^{\dagger}\Phi^{*}\delta\}$$
$$\textbf{setnotify}(m, ip, \Phi^{*}, \delta)$$
$$\{ipLn^{\dagger} \wedge ipLn^{*} \wedge ipR\Phi^{*}\delta \wedge ipNa\delta\}$$

where Φ^{*} = notify-how:$n^{*} \wedge$ notify-what:$n^{\dagger}$ and $n^{*} \subseteq$ {email, via hub} and $n^{\dagger} \subseteq$ {Cannot authenticate credentials from the identity provider}.

In the post condition the customer has been issued with a replacement credential. An audit trail, record of the data provided and process has been retained by the Identity Provider.

Customer interaction with the Service Provider is interrupted

For example internet connectivity is lost when on a train.

$$\{aO\delta \wedge aN^{\dagger}\delta \wedge ipP\delta \wedge ipR\Phi\delta \wedge ipR^{\dagger}\Phi^{*}\delta\}$$
$$\textbf{setnotify}(m, ip, \Phi^{*}, \delta)$$
$$\{ipLn^{\dagger} \wedge ipLn^{*} \wedge ipR\Phi^{*}\delta \wedge ipNa\delta\}$$

where Φ^{*} = notify-how:$n^{*} \wedge$notify-what:$n^{\dagger}$ and $n^{*} \subseteq$ {email, via hub} and $n^{\dagger} \subseteq$ {The interaction was interrupted. Please try again.}.

In the post-condition the customer has completed the access to the public service.

7.3 Synopsis of the chapter

The formalisation of the first two case studies boosted my confidence regarding the logic's expressiveness and adaptability to diverse environments and to different business models. The new challenge that this chapter raised, aimed in proving the expressiveness and richness of the logic in a context where data subjects interact with governmental services. The challenge in this context was two-fold, since it was not only the power-relations that overwhelmingly favour any government due to security reasons but the demanding business model whose requirements must be catered. The Identity Assurance Programme, a mock system inspired by a real project, offered the aforementioned challenges and was chosen as the third and final case study to verify the logic's efficiency.

The requirements for the this case study were elicited by analysing documents from discussions among the research community assisting the design of a system for the Identity Assurance Programme. The aim of the chapter was to formalise the specifications for the consent and revocation expectations of data subjects in such a system without any further ambiguities or inconsistencies emerging. An aim that was fulfilled in a satisfactory percentage. The formalisations of this case study successfully addressed requirements regarding data subjects' preferences of consent and revocation. However, I deemed appropriate to introduce the new action **request** to capture the needs of the complex business model. The specification where the data subject requests from the Identity Service to acquire their personal data from the Identity provider could have been formalised in a satisfactory manner without any further refinements. However, I decided to refine the logic with a novel action. The reason for this refinement is that the new action provides the advantage of capturing the fact that the request for acquiring the personal data is given to the Identity Service and the data subject (at least for the business model suggested so far in the project) does not conduct the Identity Provider. The formalisation without the new action would imply that the data subject had given consent to the Identity Provider to forward the data to the Identity Service.

As a synopsis, the challenge to apply the logic in three different contexts, with different business models and various consent and revocation preferences was successfully addressed. The formalisations effectively and unambiguously captured both the specifications of consent and revocation requirements and the specifications of the legacy systems existing in the environment, ensuring the richness and expressiveness of the logic. This expressiveness informed the requirements for the EnCoRe project, allowing us to help people understand what possible controls may be available to them, understand how their choices would affect and protect their privacy and enable them to express their expectations.

However, formalising the requirements necessary for offering consent and revocation controls and verifying mathematically the correctness of a system by excluding any undesirable behaviour is not enough to promise a successful implementation. Mistakes in the implementation may occur, resulting in functional problems and compromising the integrity of the system. Test methodologies can be defined based on the logic and the derived test suites can identify any erroneous behaviour, ensuring that the system's implementation will effectively correspond to the requirements formalisation. A novel testing methodology for generating tests suites in an automatic manner is presented in the next Chapter 8.

CHAPTER 8

The testing strategy

The concept of privacy is a fertile area and provides several opportunities for conducting research in diverse disciplines. Enabling data subjects to control the disclosure, use and further dissemination of their personal data in order to mitigate their privacy concerns raises several challenges. The initial step requires the development of a conceptual model that will underpin the second and most decisive step, the design of a language, able to formalise the specifications of requirements privacy controls.

In Chapter 3 a series of controls was discussed that encompass the notions of consent and revocation and a novel model was presented. In Chapter 4 the proposed model was the cornerstone for the development of a Hoare-style logic able to capture and formalise consent and revocation requirements. The logic's expressiveness and richness was proved by applying it to formalise three different case studies, namely the Employee case study in described in Chapter 5, the Biobank presented in Chapter 6 and the Identity Assurance Programme case study illustrated in Chapter 7. The data subjects expressed different consent and revocation preferences in each case study while the business models in the three contexts varied, raising different requirements every time. The successful formalisation of the totality of the data subjects' expectations in these three contexts and the efficiency with which the diverse business requirements were captured, verified the effectiveness of the logic.

However, concluding this thesis by proving that the design process of a system can achieve the proposed functionality without any ambiguities occurring would have left the purpose of this thesis incomplete. Equally important to the design process of a system is the implementation process. Errors in the implementation may instigate problematic behaviour of the system and result in misinterpretations of data subjects' consent and revocation expectations. An appropriate solution to

identify errors during the implementation process is the generation of a series of tests, to verify the integrity of the system. Since the logic can effectively capture all the requirements in a mathematical manner, it must be the basis of the testing strategy to ensure that any errors identified by the tests will be the result of a mistake in the implementation and not a misinterpretation of a requirement or an ambiguity emerged at the design process.

The purpose of this chapter is to describe a strategy for creating test suites to gather evidence of the correctness of the consent and revocation controls offered by any system. The strategy combines two procedures: Procedure 1 uses a novel formal language to elicit and document unambiguous requirements. Procedure 2 uses such formal descriptions of requirements to generate test suites. Translating the method into machine-readable language will allow the creation of automated test suites for a specific system. The strategy is demonstrated here by application to an aspect of the Employee case study. The intention is to create a testing strategy that could be applied to any system required to handle the life-cycle of consent and revocation controls imposed on data. Since there is a gap in the literature and no privacy testing methodology has been designed, this thesis extends the literature by developing a novel testing strategy which is the final contribution of this thesis to knowledge.

In what follows, the next Section 8.1 explains the value stemming from using the logic to underpin the attempt for a novel testing strategy and identifies the gap in the literature. In Section 8.2 the novel testing strategy is presented and more specifically, two procedures which generate the final testing suites are discussed. In Section 8.3 the novel testing strategy is demonstrated by applying it on the Employee case study to verify that the appropriate test suites are generated. To further verify the strategy, I adjusted it for the needs of the EnCoRe project and applied it on the first EnCoRe prototype implementation, designed for the Employee case study.The results of the testing methodology and the limitations which the strategy experiences are presented in Section 8.4. Finally, in Section 8.5 I briefly mention the key elements of this chapter.

8.1 Why formal methods in testing

The challenges raised in creating privacy test suites for a user-centric system are two-fold. They derive from the important role which the data subject has in controlling

how their personal data is handled by the system, and from the privacy issues that need to be addressed to ensure that the data will be handled in accordance with the data subject's wishes expressed in the form of consent and revocation controls.

A formal language is used to describe test requirements rather than natural language in order to provide clear and unambiguous results. This clarity is guaranteed by the existence of mathematical semantics.

In the literature there are limited references to testing privacy properties. To my knowledge the most comprehensive privacy-testing methodology is proposed by the PRIME Project [52], which championed the development of common criteria and privacy-protection profiles. They proposed core privacy properties, well-defined within the academic community, such as: anonymity, unlinkability, unobservability, undetectability, and pseudo-anonymity. However, the assessment of these attributes is still unsuccessful. There are no test suites designed to assess the effectiveness of consent and revocation controls for the handling of personal data, thus the test strategy is considered to be novel.

8.2 Presenting the testing strategy

The aim is to perform automated tests to ensure correctness of the implementation, by reference to a set of requirements derived from the case studies. The strategy will generate tests to assess functional requirements, specifically by focusing on ensuring that the consent and revocation related functions behave as expected. The tests will also be applied in the first implementation of the EnCoRe system, enabling the EnCoRe project to gain confidence in the integrity of the EnCoRe system. Proving correctness for the EnCoRe system is "elusive" [92] and in this strategy the focus is only on privacy requirements. Non-functional requirements exceed the purpose of this thesis and are not addressed*.

The strategy comprises of two novel procedures. The first one aims in eliciting testing requirements based on a formal language, while the second processes the results of the first procedure to generate a list of tests in a machine-readable format, suitable for automation within a test harness. I believe that functional testing complements the requirements formalisation [11]. The requirements have been identified

* The neglected non-functional requirements have two different sources. They derive from the assessment of the security properties, which is not relevant to privacy properties and from the complexity of privacy features, such as aggregation or anonymity, which created ambiguities in an attempt to formalise the requirements for the first EnCoRe case study [11].

and expressed using the Hoare logic described in Chapter 4. The Hoare logic is able to express all the states of a system capable of handling consent and revocation controls. Actions are given in the form of triples that describe a transition from one state to another.

The testing strategy model comprises of initial states, transitions and final states. The states are identifiable, finite in number and expressed with the Hoare logic.

According to a testing report of the British Computer Society (BCS) [130], a strategy for testing a state transition system should specify:

- The starting state
- The input to that state
- The expected output
- The expected final state

With the novel Hoare logic presented in Chapter 4 all the aforementioned attributes for a successful testing strategy can be described with clarity. The desirable initial state is captured by the pre-condition of the triple, the input that triggers the transition is defined by the action and the expected final state is described in the post-condition. Outputs from the final state, are captured with the form of obligations.

In order to clarify how the requirements are expressed, the notation used is reminded. Each action corresponds to a requirement of the following form:

$$\{\textit{pre-condition(rights/permissions)}\}$$
$$\textbf{action}(a, b, \delta)$$
$$\{\textit{post-condition(rights/permissions/obligations)}\}$$

The pre-condition comprises of rights and permissions. Every right consists of a sequence of three letters. The first letter denotes the actor that pertains the specific right, the second letter describes the nature of the right (right to process data or right to share data) and the third letter denotes the data that the right applies to. The permissions are expressed in variables which constrain specific rights.

The action describes a transition from one state of the system to another and denotes the actors who participate in this transition. The first actor is the initiator of the action and the second is the actor influenced by this transition.

The post-condition, in analogy to the pre-condition, comprises of rights and permissions. In addition, it could contain obligations, which are two-folded. They are actions that need to be triggered in the future, under certain conditions or actions that should be cascaded to third parties in order for the post-condition to be completed. In the latter case, a third actor is also influenced by the transition from one state to another.

The state of the system comprises of:

- Actors.
- Rights, predicates of the logic that are either true or false.
- A number of consent and revocation variables, that define the dimensions in which restrictions can be imposed on data use.
- The actual values of these variables.

Each action can be triggered when the pre-condition is met and when completed could either

1. Alter rights on one actor
    - Alter rights on more than one if there exists an obligation in the post-condition
2. Update data
3. Change variables
4. Set notification rules
5. Send notifications

8.2.1 Procedure 1: Eliciting test requirements

With the procedure illustrated below, requirements for the testing suites are elicited. Based on the formalisation of the system's requirements, every requirement is analysed and the derived factors and results will define the test suites on the next procedure. More specifically, the actors of the system and those that participate in the specific formalisation are identified, the rights that are altered by the action and

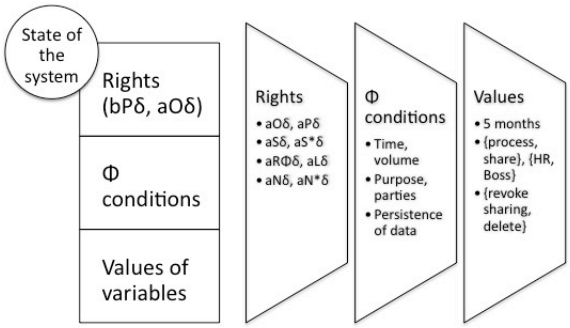


Figure 8.1: The state of the system

the values of the variants that restrict every right are clarified. Finally, any notifications or obligations that require further action from third parties, are distinguished and taken under consideration.

Below is illustrated the schema of the first model:

1. Identify the actors of the system (from the state of the system).
2. Identify actors participating in the action ($\mathbf{grant}(\mathbf{a}, \mathbf{b}, \delta)$ implies that the actors involved are a and b).
3. Identify whose actor's rights are influenced by the action ($\mathbf{grant}(\mathbf{a}, \mathbf{b}, \delta)$ implies that the actor whose rights are influenced is b).
   (a) Check for obligations in the post condition.
4. Identify the class that the action belongs to. Five different classes have been identified, namely:
   (a) **grant** actions.
   (b) **revoke, delete** actions.
   (c) **notify** actions.
   (d) **change consent and revocation variables** actions.
   (e) **update** actions.

5. Identify which rights are influenced from the action. According to the class that actions belong, rights may be added, reduced or remain the same.

6. Identify the variables of consent and revocation that are influenced and the attributed values that these have. According to the class that actions belong to, variables could be added, removed or result in a change of their values.

8.2.2 Procedure 2: Producing test suites

The factors and results identified from the previous procedure, are used as variables that influence the creation of the test suites. In order to produce test suites the following must be considered:

1. Verify that the pre-condition of the action is true.

2. Verify that only the identified actor has been influenced from the action and not the other actors of the system.

3. The rights have been altered appropriately when actions belong to the a, b and d class, data changed when actions belong to the c class and variables changed when actions belong to the e class.

4. No more/less rights have been added/reduced to/from the actors.

5. The values of the variables that constrain the rights are respected by actors.

6. Notification was sent/ not sent to the appropriate actor after the action was completed.

The test suites are designed to exercise "valid transitions" between states. Since the final state of the system has been formalised, based on Hoare rules and axioms one could test only the actions that are allowed to be triggered from that state. However, test cases may also be designed to test that "unspecified transitions" cannot be triggered [130]. This distinction allows the testing strategy to be simple by only testing the valid transitions or more thorough by verifying that transitions prohibited by the Hoare rules, are also denied by the implementation of the system.

The test suites generated for each action of the Hoare logic provide a black box of tests, meaning that every time a specific action is triggered and the transition from one state to another is completed, the tests required to validate the correctness of such a transition will remain the same.

There are though limitations to the model. For testing sequential actions 1,2,3 leading from state A of the system to state D, the system should generate tests for each action separately, testing the transition from state A to B, from B to C and from C to D.

Concurrency of actions is another limitation. When two or more actions are triggered simultaneously, or an action is triggered before the system has reacted on a previous transition, the testing suites generated by the model should be complemented with further tests to provide efficient assessment.

8.3 Applying the testing strategy to the Employee case study

Tests for all the different use cases of the scenario have been generated, but for the purpose of this thesis I will apply the strategy and develop the testing requirements for two use cases only. In the first use case, Mary has just been hired by a company X. In the second use case, Mary resigns from company X. These use cases provide complex situations and conflicting actions, since in the first use case Mary is consenting to the use and sharing of her data, whereas in the second use case, Mary requests her data to be deleted.

Testing verifies that the interactions of the system under examination with the environment through "points of control and observation" [92] conform with specifications. The consent and revocation specifications, for this particular case study, have been identified and formalised in 5. The environment of the system must be captured and according to Armando et al [21], in order to describe the testing environment one needs to define the System Under Test (SUT). The actors who use the specific EnCoRe system are defined as SUT and all the other actors are simulated by the tester. In this case study, the SUT is HR department of the X company, which is represented by the actor h in the system model. The actors Mary, Mary's boss and third parties are all simulated by the tester and comprise the environment of the system.

8.3.1 Mary is hired by company X

Before Mary is about to start in her new position she fills out various forms for the Human Resources (HR) department, including necessary health information. She signs a form agreeing to the terms and conditions which are stored by HR.

$$\{mO\delta\}$$
$$\mathbf{grant}^{\dagger}(m, h, \delta, \Phi)$$
$$\{hL\delta \wedge hP\delta \wedge hS^{*}\delta\}$$

where Φ = destination:$\Pi \wedge \neg\pi \wedge$ purpose:$p \wedge$ time duration:$t \wedge$ times processed:t^{*} and $\Pi \subseteq$ {Mary's boss} , $\pi \subseteq$ {third parties}and $p \subseteq$ {internal purposes}, $t \subseteq$ {One year - five years} and $t^{*} \subseteq$ {One time - 100 times}

The semantics of the formalisation are: m= Mary, h= HR department, δ= health information. Π is a variable that allows the HR department to share data only with Mary's boss, π restrains the HR department from sharing Mary's data with third parties while p defines the purpose for which the data should be processed. Furthermore, there are variables describing the duration of consent, t denotes the years that data should be stored for, and t^{*} the number of times the HR department may process the data.

All the actions performed in the system are defined by the system administrator and in this specific case study, the HR department. Thus, the options which Mary can choose from are pre-defined by the HR department. Furthermore, all the actions may invoke changes in the rights of the actors. In the above formalisation, before the action Mary was the owner of the data. After the action, there is a transition to a state where the HR department possesses the right to process Mary's data ($hP\delta$), the right to store Mary's data ($hL\delta$) and the right to share Mary's data($hS\delta$), all of which are restrained by conditions described in Φ.

By applying the first procedure, what is identified are the actors of the system, the rights that each actor has, the variables that restrict the rights and which rights should be altered. In essence, requirements that allow to define the initial state , the triggering action and the final state of the system are elicited.

The actors of the system are four, namely:

1. Mary
2. HR department
3. Mary's boss
4. Third party

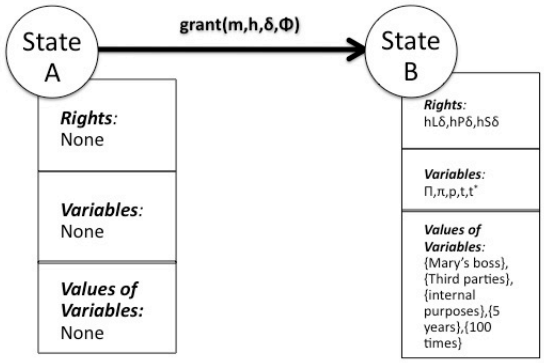


Figure 8.2: Transition from the initial state to the final state with the grant action

From these actors, those implicated into the action are Mary and the HR department. The latter actor is influenced by this action. The action is part of the "grant" class and there are no obligations created. With this action the rights that are influenced are three, namely:

1. Right to collect
2. Right to process
3. Right to share

The data pertaining to the action is Mary's health data and the variables defined are five, namely:

1. Destination (Mary's boss) and not (Third parties)
2. Purpose (internal use)
3. Time of consent (Up to five years)
4. Times data processed (up to 100 times)

These results are presented in the Table 8.1:

Procedure 2 is then applied to the results of Procedure 1, recorded in Table 8.1, and produce the test requirements described in Table 8.2.

Table 8.1: Results from the application of the first procedure

Actor	**Rights on data**	**Rights to change by the action**
Mary	owner of the data	none
HR department	none	store, process, share
Mary's boss	none	none
Third parties	none	none

It is essential to verify that only the HR department has obtained the appropriate rights and that Mary's choices are enforced. Thus, the creation of tests to verify whether the HR department has obtained the right to collect, process and share Mary's data and if her choices (variables) are enforced, is necessary. Further testing is required to examine if the HR department has obtained more rights than those mentioned and if any other actor of the system was influenced. The last tests aim at verifying that transitions prohibited in the Hoare logic are not allowed by the implementation.

Table 8.2: Testing suites generated by the application of the second procedure

Actor	**Test**	**Pre-condition**	**Post condition**
Third party	Attempt to access Mary's data as other users	No Access	No Access
Mary's boss	Attempt to access Mary's data held within the HR department	No Access	No Access
HR department	Attempt to process Mary's data for 99th time	No Access	Access Granted
HR department	Attempt to process Mary's data for 101th time	No Access	No Access
HR department	Attempt to process Mary's data after five years	No Access	No Access
HR department	Attempt to share Mary's data with Mary's boss	No Access	Access Granted
HR department	Attempt to notify Mary	No Access	No access
HR department	Attempt to release Mary's data to third parties	No Access	No Access

Each row effectively describes a single test of the system: a test harness must first establish a correct system state (satisfying the pre-condition), attempt the access specified, and then observe whether the resulting system state meets the postcondition. If it does, the test is considered to be passed; if not, a failure is registered. Of course, given all the unconstrained factors that contribute to the system state, there is no guarantee that this result is completely determined; the same test might give the opposite result in other circumstances. The role of testing, however, is to contribute to evidence-gathering about the correctness of any implementation, not to be the sole arbiter, and tests passed will contribute to confidence.

The testing suites are executed once in every transition. Firstly, it is tested if the triggering of the action from the initial state was valid. The expected result for the tests is described in the third column of the Table 8.2 above. The second set of tests aim to verify that the transition has resulted in reaching the desirable final state and the result for each test is defined in the fourth column of the Table 8.2.

8.3.2 Mary leaves the company

Mary decides to leave the company. She wishes to revoke her consent regarding the use of her data and requires all data to be deleted.

$$\{mO\delta \wedge hL\delta \wedge hP\delta \wedge hS\delta \wedge hR\delta\Phi\}$$
$$\mathbf{delete}(m, h, \delta)$$
$$\{\neg hL\delta \wedge \neg hP\delta \wedge \neg hS\delta\}$$

When the transition is completed, in the final state of the system the only actor with rights should be Mary.

The semantics of the formalisation are: m= Mary, h= HR department, δ= health information. $hR\delta\Phi$ is a right denoting that the HR department respect the choices (Φ variables) which Mary gave in the past and restrict the process and sharing of her data.

By applying the first procedure, the information presented below is gathered:

The actors of the system are four, namely:

1. Mary
2. HR department

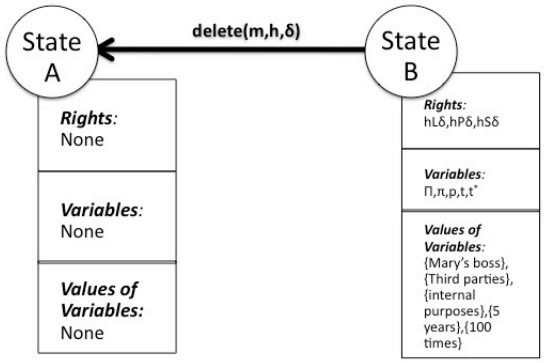


Figure 8.3: Transition from the initial state to the final state with the delete action

3. Mary's boss

4. Third party

From these actors, those implicated into the action are Mary and the HR department.The latter actor is influenced by this action. The action is part of the "revoke" class, meaning that the rights are removed and there are no obligations created. With this action the rights that are removed from the HR department are three, namely:

1. Right to collect

2. Right to process

3. Right to share

The data pertaining to the action is Mary's health data and the variables which were defined by Mary in previous transitions are five, namely:

1. Destination (Mary's boss) and not (Third parties)

2. Purpose (internal use)

3. Time of consent (Up to five years)

4. Times data processed (up to 100 times)

Table 8.3: Results from the application of the second procedure

Actor	**Rights on data**	**Rights to change by the action**
Mary	owner of the data	owner of the data
HR department	store, process, share	none
Mary's boss	none	none
Third parties	none	none

These results are presented in Table 8.3.

Procedure 2 is then applied to the results of Procedure 1, recorded in Table 8.3, and produce the test requirements described in Table 8.4.

It is important to verify that the appropriate rights were revoked by the HR department and that Mary's data is deleted. Thus, the creation of tests to verify whether the HR department still possesses the right to collect, process and share Mary's data is essential. Further testing is required to examine if the HR department has any other rights and if any other actor of the system was influenced.

Table 8.4: Testing suites generated by the application of the second procedure

Actor	**Test**	**Pre-condition**	**Post condition**
Third party	Attempt to access Mary's data as other users	No Access	No Access
Mary's boss	Attempt to access Mary's data held within the HR department	No Access	No Access
HR department	Attempt to process Mary's data for 99th time	Access Granted	No Access
HR department	Attempt to process Mary's data for 101th time	No Access	No Access
HR department	Attempt to process Mary's data after five years	No Access	No Access
HR department	Attempt to share Mary's data with Mary's boss	Access Granted	No Access
HR department	Attempt to notify Mary	No Access	No access
HR department	Attempt to release Mary's data to third parties	No Access	No Access

The test suites designed for this formalisation, and presented in Table 8.4, are the same with the previous use case, but the results of the tests are different. The testing is successful because the actions described transitions from state A to state B and backwards. The transitions were only between these two states of the system and the tests should remain the same, since the pre-condition for the one transition was the post-condition of the other and vice versa.

8.4 Results from the application of the strategy on the EnCoRe prototype implementation

The testing strategy was further applied on the EnCoRe prototype implementation, serving a dual purpose. The first aim was to further validate the testing strategy on a real system, able to capture data subjects' consent and revocation expectations and the second was EnCoRe-related and aimed to validate the integrity of the first EnCoRe prototype implementation. I adjusted the strategy to the EnCoRe specifications and matched the two procedures with the components that comprise the EnCoRe's technical architecture. The explanation of the process for configuring the testing methodology can be found in the next Section 8.4.1. In addition, I designed a pseudo-code, allowing the automated generation of tests for every action defined in the logic and implemented in the EnCoRe system.

The pseudocode was implemented by HW Communications Ltd, the EnCoRe partners responsible for the implementation of the EnCoRe system and without their valuable help the results presented in this thesis would not have been achieved. I spent several days in Lancaster, assisting in the development of the testing suites and ensuring that the tests generated with the logic, will be the same with those generated in the system.

The result was a successful generation of testing suites, adequate to address the system's integrity for the Employee case study. Although the implementation of the EnCoRe system effectively responded to the majority of the automatically generated tests, a misinterpretation of the technical architecture was revealed. More specifically, both in the technical architecture and in the logic, when the data subjects extend their initial consent, their initial preferences are not revoked but these are still enforced in the system. In the implementation, the initial preferences were deleted and only the recent consent was tracked and enforced. The problem occurred because the action of granting any type of consent initiated a format in the

data base storing the consent and revocation preferences. Slight modifications were deemed appropriate to fill the gap between the architecture and the implementation.

8.4.1 Mapping the "logic-derived tests" to the EnCoRe architecture

In the logic of consent and revocation, there is a series of actions that add and remove rights to the state of the system. Although there are several "rights" defined in the logic {L "rights" (encoding knowledge of location) and R "rights" (encoding commitment to respect conditions on specific data}, for the testing the focus will only be on three different rights namely:

- Right to Process data δ (aPδ : "a" has the right to process δ)
- Right to Share data one step further (aSδ :"a" has the right to share δ one step further)
- Right to share data transitively (a$S^*\delta$: "a" has the right to share δ "transitively", i.e. each recipient then has the right to share it onwards in subject to the same restrictions so there may be multiple hops)

By performing the action grant(a,b,δ), a gives b the right to process δ. By performing the action grant^1(a,b,δ), a gives b the right to (process and) share δ with anyone one-step further. When b shares δ with c, say, then c thereby gets the right (bPδ) to process δ (not also to share δ). In the case of grant*(a,b,δ), c would additionally get the right to share δ either one-step further (if b performed grant1(b,c,δ)) or transitively (if b performed grant*(b,c,δ)). If b decides to share δ further then the third party obtains the right (bPδ) to process δ, and either the right (bSδ) to share δ one step further or the right (bS$^*\delta$) to share it transitively, according to whether b performed $grant^1(b, c, \delta)$ or $grant^*(b, c, \delta)$. All the rights may be constrained by Φ conditions. These conditions that are described in detail in the logic, are mainly expressing data subject's choices and could create obligations. (Process δ for a specific time, delete after specific time, share only with party c, etc. . .)

System design

The system design is based on access control policies (ACPs), access requests (ARs) and obligations (Obs). ACP policies comprise Policy Decision Point (PDP)

policies and Data Registry (DR) policies. A PDP policy is an overarching policy for the organisation, including whether or not to refer to the Data Subject's policy, which is stored at the DR. An obligation is created from an ACS and could be a time-event obligation or a notification obligation. Access requests may occur in the system and their result should be as dictated by the ACPs. A result of an access request could be either "Allow" or "Deny".

The ACPs, Obs and ARs comprise a number of attributes. Some of these attributes needed to be defined for the testing. Others are 'wild cards', meaning that all possible values should be used in turn. Other attributes (marked Do_not_Care) do not influence the tests, so each can take an arbitrary value.

Mapping from logic to system design

The three actions described in the logic each create an ACP for the PDP and if there are Φ conditions then they create an ACP for the DR as well. The number of ACPs the action will create depends on the number of rights that are added.

Every single right creates an ACP. If the purpose is identified in the Φ condition then that ACP should be written as many times as there are different purposes, in order to have a unique ACP for each purpose. Thus, the only attribute that will differ among these policies will be the purpose (aside from any wild-card attributes).

The actions may also include obligations in the post-conditions. If there is a notification request or a time-request defined in the Φ condition, then in addition to the ACP, an Ob is created.

Data controller shares further

In the case where data controller "b", say, decides to share the data further (triggering one of the actions grant(b,c,δ), grant^1(b,c,δ) or grant*(b,c,δ)) then an AR is created at "b's" system and an ACP is created at "c's" system. The rules described above dictate the creation of the ACPs in "c's" system as well. It can be argued that:

- the grant(a,b,δ) action cannot lead to an action that can create ACPs for third parties.
- the grant^1(a,b,δ) action may lead to the creation of ACPs in "c's" system similar to those created by the action grant(a,b,δ) on "b's" system.
- the grant*(b,c,δ) action may lead to the creation of ACPs in "c's" system

similar to those created by the grant^{1}(a,b,δ) or grant*(a,b,δ) in "b's" system (since "b" is able to share data either one step further or transitively).

All the tests will be executed as ARs and the attributes of these ARs will be informed by the testing strategy presented in Section 8.3. Specifically:

1. The testing strategy involves deriving a set of tests from the logic of the form:
   - Actor of the action.
   - The action under test.
   - Desirable response before the action (Access-Deny).
   - Desirable response after the action (Access-Deny).
2. The "logic-derived" tests will be translated to "architecture-level" tests in the form of ARs by:
   - Checking that the system is in the desirable state.
   - Apply AR to test the state of the system.
   - Creating an ACP for the action and defining the attributes of the ACP according to the Φ conditions and the predicates of the post condition of the action (as described above).
3. The ARs will be submitted to the system and in each case the system response will be compared to the required response (Accept or Deny).

8.4.2 Limitations to the testing strategy

The novel testing strategy, although successfully applied to the EnCoRe system, has limitations. Since the testing strategy has been used to devise test suites adjusted to the EnCoRe system for one of the three case studies, some validation of its utility can be claimed. However, only a small subset of the test suite has been deployed to-date, which means that there is not enough evidence to understand the appropriateness of the strategy. Of particular concern is whether there have been tested enough components of the system, or whether instead there will remain many "dark corners" left to be explored. At the time of writing the project has not been able to deploy the strategy in earnest; some components of the system formalised by the logic and referred in the technical architecture have not been implemented yet. As a result the notification processes, the obligation actions and the actions resulting

in sharing data with other data controllers have not been tested yet. However, it is the intention of project partners in case of an EnCoRe sequel to continue developing this method further.

In addition, the strategy inherits the limitations deriving from the nature of testing. Some of the tests might have been successfully passed by the system due to the factor of luck, resulting in misleading conclusions. Furthermore, the effect of concurrent actions has not been considered in the testing strategy. The testing harness cannot probe state resulting from concurrency. In addition, there are situations where the concurrency model presented in Section 4.2.1, although it provides the desirable answer, this answer can be reached with different ways. For example, an update and an action to share data further, if executed concurrently will result in a final state where the third party will have obtained the updated data. However, according to the execution sequence, the third party may or may not acquire temporally the old data. Thus, there is an opportunity for future work to evolve it to probe such circumstances.

8.5 Synopsis of the chapter

Following the design and the successful implementation of the logic to the three different case studies, this chapter explored an auxiliary use of the logic, which completes the purpose of this thesis. Since the effective formalisation of requirements for consent and revocation cannot ensure the correct implementation of a system offering data subjects controls to mitigate privacy concerns and endangering its integrity, the final step to provide a holistic solution to the problem that this thesis sought to address, was to design a testing strategy, agnostic to technical implementation, that will generate testing suites.

The novel strategy presented in this chapter comprises of two procedures. The first procedure, based on the formalisations of the logic, elicits the requirements for the tests suites, while the second procedure, uses the results of the first procedure as inputs and generates the appropriate testing suites. The strategy was applied on the Employee case study and generated tests , verifying the testing methodology and estimating that any fail in the testing process will possible be a result of an erroneous implementation. To further verify the testing strategy, the methodology was adjusted to the needs of the first EnCoRe prototype implementation. The majority of the tests were successfully implemented by HW Communications

Ltd, the company responsible for the implementation of EnCoRe. The application of the strategy revealed a "misinterpretation" of the technical architecture in the implementation process and highlighted the need for slight modifications.

CHAPTER 9

Conclusions and future work

There have existed times in history when technological developments crafted the way people communicated, causing changes in societal norms in such a manner that the world was a different place afterwards. The advent of the Internet and its reception from our society has not been an exception as it has opened novel channels for communication, influencing the social, business and economical aspects of our lives; people can obtain access to services and products online, participate to social networks or conduct their businesses on the web. Therefore, individuals are disclosing more personal data than ever, while enterprises have the potential to store, process and share vast amount of data.

However, exploiting the benefits offered via the internet comes with compromise. The control which data subjects might have over their personal information is eroded, raising privacy concerns regarding the handling of their data. The implications of such lack of controls are profound and have been manifested with the increasing number of incidents where data has been lost, mistreated or shared further without the data subjects' knowledge. Thus, privacy as a research topic has gained in popularity in recent years, since the need to provide users with mechanisms enabling them to control the storage, use and dissemination of personal data is more imminent than ever.

The importance of privacy has been advocated in the majority of literature from scholars in a wide range of research interests, arguing that if we do not act upon the contemporary privacy threats, we will remember the privacy right with a sense of nostalgia [236]. The most prominent perception of privacy construes it as a right whose purpose is to protect individuality and exclude people from social obligations. Its value stems from safeguarding human dignity, enhancing individual creativity and catering for positions opposing the dominant [82; 45; 41; 101]. Framing privacy

to individualistic terms, although catering for valuable elements, provokes criticism due to the emerging conflicts with other primary rights and societal interests. Privacy is considered to threaten societal solidarity [86; 119], as it is deemed antagonistic to civil liberties, human rights and freedom of speech [274], while hindering social surveillance by concealing malicious activities [243; 212]. In addition, feminists dispute privacy as a right claiming that it alters the authoritative relations between sexes in favour of men, concealing domestic violence [166; 236]. Others attach an economic perspective to their criticism, suggesting that privacy hinders the disclosure of information critical for businesses to operate in a free market, resulting in a loss of trust and sharing of discredible information[211; 85].

This thesis, has adopted a different approach to privacy and perceived it as a right that has social value whose importance stems not only from the individualistic benefits but from the benefits it offers to the common good and the prosperity of society. Loss of privacy results in loss of freedom, openness and transparency in social life and since privacy is critical for the protection of a plethora of interests, its value should be assessed depending on the issue that is addressed each time [236; 44]. These values should be balanced against the value of contradicting interests and the optimal solution for the society must prevail. The perception of privacy adopted in this thesis implies a balancing approach which was identified and effectively addressed.

I decided to construct and address privacy concerns as a problem of achieving a balance between individuals' benefits and the public good. As several examples in history have demonstrated, legislation endeavours to establish functions to balance individuals' right to privacy and the common good. However, the use of technological innovations alters the power relations and creates a constant need for re-establishing the balance. It is impossible for the legislators to anticipate technological developments and eliminate privacy concerns which emerge from this vicious circle, thus alternatives must be sought. The research presented in this thesis endeavoured to maximise the benefits both for the individual and the public good, while effectively addressing privacy concerns.

By focusing on the concepts of consent and revocation, this thesis has aimed to mitigate the problem of privacy concerns by developing a logic able to capture data subjects' expectations to control the flow of their personal data. Since privacy is a fertile area for conducting research, the successful solution presented in this thesis contributed to knowledge in multiple ways. The next section discusses the core novel aspects of this thesis.

9.1 Contributions to knowledge

The goals of this thesis were achieved by following a demanding and rarely adopted approach as it combined elements from both social and computer sciences disciplines, offering an interesting blend. The combined approach created opportunities for contribution to various layers. Departing from the literature, the first aim was to obtain a holistic understanding of privacy. To that end, I classified the literature in privacy conceptions and identified their limitations. The most dominant conception in the literature, conceives privacy as controls and a further examination of the controls identified by researchers was conducted, resulting an observation of a lack of more dynamic controls able to capture the volatile nature of the online environment. In addition, the most prominent means to provide these controls was the concept of consent. Once again, studying the literature on consent revealed a gap, since common practice treats the process of giving consent as an one-off event, remaining oblivious to the notion of revocation that can, as proven in the thesis, introduce a dynamic element. In addition, recent legislative efforts were presented and provided evidence for the legal need of a system able to handle consent and revocation controls, since the European Union has shifted the focus from the right to be let alone, to the right to be forgotten [80].

The gap in the literature on revocation and the legal evidence for developing such a concept, was the basis for the first contribution to knowledge achieved in the thesis. Based on the literature presented in Chapter 2 and on the analysis of data generated by focus groups, I developed a conceptual model of consent and revocation and coined the term of "informed revocation" in Chapter 3. The focus groups, were conducted for the needs of the EnCoRe project and were designed by the team of the London School of Economics, led by Dr. Whitley, to whom I am most grateful for allowing me to have access on the transcripts of the focus groups. It should be noted that I participated in shaping the questions posed to the participants of the focus groups, since those were decided in the general EnCoRe meetings.

To analyse the data in a systematic and rigorous manner, elements from qualitative research were embraced. The content analysis approach was considered the optimal methodology to serve the purpose and the generated data was categorised based on the literature on consent. The analysis revealed a new dimension for consent which was implicit in some researchers' treatment [47] and identified different types of revocation resulting in a novel conceptual model of consent and revocation. The model details the different kinds of control that users desire to

exercise over personal data concerning them that is held by an enterprise. Furthermore, a phenomenon was identified where the data subjects altered their choice of revocation when informed of the many different kinds that exist. The term of informed revocation was introduced to capture this change of behaviour. I also argued that the notion of informed revocation, although coined in analogy to Faden and Beauchamp's [89] term of "informed consent", does not inherit its limitations, rather assists on addressing parts of its criticism.

The second contribution of this thesis involved elements from the computer science area. The study of the stream of literature concerning formal methods which address privacy issues, revealed a lack of languages that focus on formalising the concepts of consent and revocation. So far, approaches to formal methods for privacy management were either high-level, relying on legal compliance, risk assessments and addressing privacy in already implemented systems (such as Barth et al [26]), or low-level, focusing on technical implementation and often disregarding important legal and business considerations. A characteristic example of a low-level approach is P3P [257], which translates privacy policies written in natural language into machine-readable formats. There was, however, a gap between implementing various policies at the high level, and being able to provide traceability into the low-level policy implementations. Thus, this thesis contributed to a niche space in the field of formal methods and their use, since none of these attempts tried to address the issues of consent, revocation and deletion of data.

I designed and presented a novel Hoare-style logic, providing the means for reasoning about consent and revocation controls, via which data subjects can express their privacy preferences, and formalising the specifications of such processes in a system. Controls for consent are defined as data subjects' privacy preferences pertaining to specific personal data, while revocation controls are defined as the process that corresponds to the withdrawal of consent. The logic comprises a set of rights for principals in the system, a set of actions that express a transition in the system from one state to another and Φ conditions which set the values for consent and revocation variables. The rules of the logic are given in the form of triples, where the pre-condition containing rights and factors defines the initial state, the action describes the activity in the system and the post-condition defines the concluding state of the specific activity and constitutes of rights, facts and possibly obligations. Obligations define sequential to the activity actions which must be performed in the system either to complete the initial action or as a consequence of it.

I have provided an exhaustive list of rights and actions that successfully cap-

tured all the specifications not only for the data subjects' consent and revocation expectations, but for the business and legal requirements as well. Regarding the Φ conditions comprising of consent and revocation variables, I deliberately attached to those restrictions an elasticity, allowing the creation of new variables that can effectively capture diversities regarding the contexts where the logic is applied on. The logic is designed to satisfy healthiness conditions, defined in such a manner which verified that every valid behaviour described in the logic can be achieved, while ensuring that undesirable behaviour is prohibited. The use of the Maude software was deemed necessary to construe proofs for the validity of the healthiness conditions. Due to the nature of the Hoare-style logic, where actions are considered to be sequential, the need for a concurrent model was acknowledged, since it is highly unlikely that all the actions in a system will be executed in a sequential manner. Problems occurring due to concurrency reasons were identified and a first approach to solve those was proposed.

The logic was not designed in a single attempt, rather it was subjected to a gradual refinement. The case study methodology was preferred to underpin the refinement process and to validate the final form of the logic presented in Chapter 4, in terms of richness and expressiveness. The case studies where the logic was applied, albeit designed for the needs of the EnCoRe project, were deemed appropriate to serve the purpose of validation, because they capture the three different kinds of power asymmetries described in Chapter 3. When adopting a data subjects' perspective, there exist three different contexts where the power relations between the data controller and the data subject differ in such a degree that the privacy expectations in each of these environments vary. The three case studies illustrated in this thesis, captured the power asymmetries of the three different contexts, in an attempt to identify and formalise as many different privacy expectations and business models as possible.

The Employee case study was the pilot scenario and the application of the logic revealed ambiguities which were categorised into two kinds, according to the source of their existence. Furthermore, these ambiguities were addressed and their solutions resulted in the final version of the logic. In Chapter 5, the application of the final version of the logic served a different purpose, as the aim was to verify the richness of the logic by effectively formalising all the requirements elicited for the specific case study, in an unequivocal manner and without any further ambiguities emerging. The adequate formalisation of all the use cases provided evidence for these claims.

The second case study described in Chapter 6 raised further challenges and its

aim was to question the ability of the logic to capture requirements for a different context, with different consent and revocation expectations that also involved legacy systems in place whose requirements must be catered for. The Biobank, was the perfect candidate as the treatment of medical personal data requires specific privacy preferences and the presence of the Biobank's researchers ensures a demanding audience arguing for specific business requirements. In order to elicit the necessary requirements, content analysis was used to analyse transcripts of focus groups, specifically designed to identify and capture patients' and researchers' expectations of a system operating in the Biobank.

The logic was successfully applied in the new context and was used to formalise all the new privacy preferences and business requirements with minor adaptations. Those adaptations were the introduction of new variables that enhanced the richness of the logic. A subsequent number of formalisations with various different examples of consent and revocation controls provided evidence of the successful application of the logic to the Biobank case study. I illustrated how diverse mechanisms of consent and revocation may allow donors of the Biobank to express their preferences and acquire control of their samples and data. Furthermore, with the formalisation of the requirements for such a system, I effectively validated that patient choices are unambiguously described in order to be translated into privacy policies and enforced into the system.

The third and final case study, described in Chapter 7, considered the power asymmetries when the data subject interacts with governmental services. The Identity Assurance Programme, a mock system inspired by a real project, provided all the necessary situations to reveal the various privacy preferences which data subjects expect to be allowed to request, as well as a novel business model which raised new challenges. The requirements were elicited by analysing the documents provided by various researchers, that formed the team assisting the implementation of the project.

All the consent and revocation requirements were successfully formalised without any further ambiguities emerging, boosting my confidence regarding the richness of the logic and its applicability in diverse environments with minor adaptations. However, the novel business model required the slight modification of one action, resulting in the introduction of a new action to the logic. The new action captures the request of a third party, on behalf of the data subject, for acquiring access to personal data stored in a data controller's database. The formalisation of the three case studies verified that the novel logic is general applicable and the emerging

ambiguities are minor and solved without subsequent refinements.

The final contribution of this thesis to knowledge is a novel testing strategy which provides evidence that the implementation of a system designed to capture the management of consent and revocation controls is valid. The stream of literature studying testing strategies for privacy-enhanced systems is still in its infancy and little has been achieved so far regarding the development of a methodology for testing privacy properties. Thus, this thesis also filled a gap toward this direction.

The proposed model in Chapter 8 comprised of two novel procedures designed based on the logic of consent and revocation. These procedures develop automated testing suites which effectively validate the correctness of consent and revocation controls, while remaining agnostic to the technical architecture of the system. The strategy was applied on the Employee case study to prove its practicality, at least as far as individual actions were concerned.

In addition, I configured the procedures to match the technical architecture of the EnCoRe prototype implementation designed for the Employee case study and provided a pseudo-code to automate the process of generating testing suites. The pseudo-code was implemented by the HW Communications Ltd, the EnCoRe partners responsible for the implementation of the EnCoRe system, to whom I am most grateful for their assistance. The aim of applying the testing strategy to the real system was two fold. To further validate the applicability of the strategy and to test the correctness of the EnCoRe system. Since the logic was proven valid, any failures in the testing process would have been the result of a faulty implementation. The testing suites were successfully generated, testing several activities in the EnCoRe system. The strategy produced a number of tests in every occasion and resulted in identifying a misinterpretation of the technical architecture in the implementation of the EnCoRe's database.

9.2 Contributions to the EnCoRe Project and Dissemination of Project Results

The logic had a protagonist role for the EnCoRe system, ensuring that everything the EnCoRe prototype implementation offers can be achieved without further ambiguities and undesired behaviour. The EnCoRe project ran to conclusion on June 2012 and is the pioneer of a system enabling users to express their privacy expectations by delivering a wide range of technology and procedural controls designed to

provide data subjects with consent and revocation life-cycle management over their personal data.

The solutions offered by the project provide novel technical, legal and compliance aspects. A number of core concepts have been defined and classified and the relationships and linkages between these concepts have been identified creating a common language for people to use, in order to be on common grounds when the issue raised is privacy controls.

Regarding the technical architecture, a composure of diverse elements, both software and service components, capture the data flows between data subjects, data controllers and other technical systems, cater for compliance and regulatory environments while all the technical procedures necessary to manage and enforce data subjects' privacy expectations are encompassed. The implementation of the EnCoRe technical architecture, resulted in user interfaces offering consent and revocation options and notification capabilities to inform data subjects of their data handling.

Drawing on legal and regulatory aspects, the EnCoRe consortium commissioned a position document, reviewing the new EU General Data Protection Regulation, to the public consultation that the Minister of Justice in the United Kingdom organised to garner view of interested parties [50]. In addition, departing from the latest social science research, a document describing a compendium of how consent and revocation is exercised in current business and research practices was published.

Another important contribution that the EnCoRe project achieved, was a novel compliance framework able to provide evidence for assessing a system's standards when offering consent and revocation controls. The project innovated compliance tools to support evidence demonstrating that data subjects' choices are enforced within the system. Amongst various periodic and time-bound or continuous designed tools, test suites based on the logic are some of the most important elements of the compliance framework. In addition, the evidence gathered from the compliance process provides input for the novel privacy risk assessment methodology.

9.3 Future work

The research findings presented in this thesis provide opportunities for various paths of future work either by drawing upon and addressing the limitations of the thesis, or by extending the use of the key contributions, assisting in the growth of the

research field. The first opportunity for future research derives from addressing limitations of the existing logic. Although the logic provided a mechanism to formalise requirements and illustrate proof of correctness for systems managing consent and revocation controls, it could still benefit from advancements in the concurrency model and the automated generation of testing suites. In particular, its concurrency model does not account for time delays that will inevitably exist in an implementation of a real system. If users are able to make rapid changes in their choices, then delays in the system may mean that the system handles data according to out-of-date choices, so it would be unreasonable to require instantaneous changes in how data is handled. A first approach was presented in Chapter 4 but I believe that this can be refined by further work, perhaps by explicitly allowing for system delays with the use of recent snapshots. A first approach is presented in an EnCoRe public deliverable [6].

Limitations on the concurrency model have also an impact on the testing strategy. A key research development could be to consider the effects of concurrency and how the test strategy should evolve to probe them. Further application of the testing method to other case studies can assist in a detailed assessment of the applicability of the approach. In addition, the method could be complemented with other compliance mechanisms, such as run-time monitors, to provide an holistic compliance approach [6]. To-date, systems that provide similar functionalities to EnCoRe have not been designed yet. However, if in the future a similar system exists it could provide an opportunity to verify my intention to design the testing strategy in a manner agnostic to the technical architecture and implementation.

The logic merely addresses the issue of aggregation by restraining the purpose of processing data. I do not claim to have solved the issue thoroughly, due to the fact that data controllers may derive new data from processing and link it with data publicly available. Future research could focus on imposing consent and revocation restrictions on data that is derived from aggregation. Embracing elements from other variations of logics to include, for example epistemic and temporal modalities, may further assist in addressing the problem of aggregation. Additionally, temporal modalities may enable the formalisation of time and define the Φ conditions in a mathematical form. Furthermore, future research focusing on these modalities could provide the foundation to formalise models that attribute liability and ownership of data, providing concrete solutions to the conflicting requirements of anonymization and traceability. Establishing compliance to privacy requirements is an elusive but necessary task under the threat of either financial penalties posed by

privacy regulators or reputation damage caused when consumers' trust evaporates. Thus, the attribution of liability when privacy breaches occur, proof of conformity to legislation regarding data protection and reasoning about data ownership provide opportunities to mitigate risk and ensure compliance with regulation and best practices.

A different approach to future work could draw upon the achievements of this thesis and elaborate further. There exists a niche space in the field of formal methods to contribute to the comprehension of the complexities in the relationships between users and organisations. In the logic, an assumption expressed by the "$bR^{\dagger}\Phi\delta$" right suggests that the data subject has reached an agreement with the data controller regarding the available controls. Conflicts in privacy requirements emerging from different stakeholder interests should be further examined with the aim to provide a refinement of high-level policies. Further research could focus on developing a conceptual model for data usage policies that will take into consideration diverse privacy requirements from users, as well as legal, business and technical stakeholders. The logic can be used to translate privacy policies and form a strategy to resolve policy conflicts. Towards this aim, embracing other languages such as SecPal [30], can assist in reasoning about policy refinement and trade-offs between stakeholders.

In addition, while this thesis has considered the perspective of the user, another direction of investigation is to adopt an enterprise's perspective and form a model of business prerequisites before engaging a system enabling users to control and acquire access to their personal data. There are certainly issues that organisations, no matter how willing to implement a system catering for users' privacy, will face. Recent legislative efforts may have provided solid grounds for the existence of such a system, but as denoted in Chapter 2 the implementation challenges, the restraints and the changes in the business ordinary practices that organisation will have to undertake are still issues to be addressed.

Regarding the implementation challenges, organisations will be required to function in a privacy-friendly mode, storing the minimal amount of data required to achieve their tasks and keeping a record of how and to whom data is processed and shared. Moreover, the concept of revocation raises significantly the level of difficulty, since business may be obliged to delete data or revoke permissions that are vital for their economic growth. In a similar vein, the process of revocation may hinter current business functions and contradict with contemporary business practices. For example, the adoption of cloud computing to increase business agility and reduce liability will be in conflict with individuals' request to control the sharing of

data to third parties. In addition, cross-boundary sharing of data, which is common practice within cloud providers, opposes privacy requirements that a system offering consent and revocation controls should consider. This thesis describes the full range of revocation controls. There will be circumstances, however, where some of these controls will be impractical to implement. Therefore, a comparison of the two approaches, developed by focusing on different perspectives, may shed light in the conflict resolution methodology and bring to surface problems that will occur from the implementation of a system offering consent and revocation controls.

Returning to the philosophical question of balance described in Chapter 1, it is evident that a system offering consent and revocation controls cannot ensure that the balance in society will not be disturbed. It is not the belief of the author that such a technology could be achieved. Rather, offering consent and revocation options to the users provide the means to facilitate the process of achieving balance. The issues described in Section 1.1, challenging the value of privacy are still present in the consent and revocation model. The consent and revocation controls, however, may provide a fertile area for communication and common understanding between users and organisations. Thus, enterprises could better understand the concerns of their customers, whereas individuals will be better informed why in some cases their privacy requirements cannot be fully implemented. Providing a platform for raising awareness regarding privacy requirements and facilitating the process of negotiation between enterprises and users is the necessary step to mitigate problems that disrupt the balance.

Implementing consent and revocation controls raises technological, legal and business challenges. Thus, researchers need to combine effectively diverse scientific fields that are not necessarily complementary. Developing a logic based on a conceptual model of consent and revocation, applying it on real-world scenarios to identify and address emerging difficulties, and providing a testing strategy underpinned by the same logic, is the first decisive step towards this objective.

APPENDIX A

Code for the implementation of proofs in Maude

This chapter provides the code used to capture the state-model illustrated in Chapter 4 that was used to presents the proofs of the healthiness conditions described in Section 4.3.

The code is:

```
mod SIMPLE-4 is
        sort State .
        var W : Bool .
        var M : State .
        ops O L P S S1 T X : -> Bool [ctor] .
        op a : -> State [ctor] .
        op b : -> State [ctor] .
        op c : -> State [ctor] .
        op d : -> State [ctor] .
        op e : -> State [ctor] .
        op f : -> State [ctor] .
        op g : -> State [ctor] .
        op h : -> State [ctor] .
        op i : -> State [ctor] .
        op j : -> State [ctor] .
        op rights : State -> Bool [ctor] .
        eq rights( a ) = O .
        eq rights( b ) = O and L and P .
        eq rights( c ) = O and L and P and S  .
        eq rights( d ) = O and L and P and S and S1 .
        eq rights( e ) = O and L and P and T and X .
        eq rights( f ) = O and L and P and S and T and X .
        eq rights( g ) = O and L and P and S and S1 and T and X .
        eq rights( h ) = O and L and T and X .
        eq rights( i ) = O and T and X .
        eq rights( j ) = O and L .
        rl [grant] : a => b .
        rl [grant] : b => b .
        rl [grant] : c => c .
        rl [grant] : d => d .
        rl [grant] : e => e .
        rl [grant] : f => f .
        rl [grant] : g => g .
        rl [grant] : h => e .
        rl [grant] : i => e .
        rl [grant] : j => b .
        rl [grantshare] : a => c .
        rl [grantshare] : b => c .
        rl [grantshare] : c => c .
        rl [grantshare] : d => d .
        rl [grantshare] : e => f .
        rl [grantshare] : f => f .
        rl [grantshare] : g => g .
        rl [grantshare] : h => f .
        rl [grantshare] : i => f .
        rl [grantshare] : j => c .
        rl [grantsharetr] : a => d .
        rl [grantsharetr] : b => d .
        rl [grantsharetr] : c => d .
        rl [grantsharetr] : d => d .
```

```
        rl [grantsharetr] : e => g .
        rl [grantsharetr] : f => g .
        rl [grantsharetr] : g => g .
        rl [grantsharetr] : h => g .
        rl [grantsharetr] : i => g .
        rl [grantsharetr] : j => d .
        rl [grantthird] : c => f .
        rl [grantthird] : d => g .
        rl [grantthird] : f => f .
        rl [grantthird] : g => g .
        rl [delete] : b => a .
        rl [delete] : c => a .
        rl [delete] : d => a .
        rl [delete] : e => i .
        rl [delete] : f => i .
        rl [delete] : g => i .
        rl [delete] : h => i .
        rl [revoke] : b => j .
        rl [revoke] : c => j .
        rl [revoke] : d => j .
        rl [revoke] : e => h .
        rl [revoke] : f => h .
        rl [revoke] : g => h .
        rl [deletethird] : e => b .
        rl [deletethird] : f => c .
        rl [deletethird] : g => d .
        rl [deletethird] : h => j .
        rl [deletethird] : i => a .
        rl [revokesharetr] : d => c .
        rl [revokesharetr] : g => f .
        rl [revokeshare] : c => b .
        rl [revokeshare] : d => b .
        rl [revokeshare] : f => e .
        rl [revokeshare] : g => e .
        rl [process] : b => b .
        rl [process] : c => c .
        rl [process] : d => d .
        rl [process] : e => e .
        rl [process] : f => f .
        rl [process] : g => g .
endm
```

Bibliography

[1] http://www.networkworld.com/news/2012/012612-privacy-scandals-255357.html?hpg1=bn, 2012.

[2] EU Directive 95/46/EC. Directive 95/46/ec of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML, 1995.

[3] A. Acquisti. Nudging privacy: The behavioral economics of personal information. *Security & Privacy, IEEE*, 7(6):82–85, 2009.

[4] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Privacy enhancing technologies*, pages 36–58. Springer, 2006.

[5] A. Acquisti and J. Grossklags. Privacy attitudes and privacy behavior. *Economics of information security*, pages 165–178, 2004.

[6] I. Agrafiotis, P. Bramhall, S Creese, M. Goldsmith, L. Hooper, P. Hopkins, K. Hughes, N Moffat, and S. Pearson. Encore compliance framework. Technical report, Technical Report D5.1, Computer Science Department of Oxford, Oxford, 2012.

[7] I. Agrafiotis, S. Creese, and M. Goldsmith. Developing a strategy for automated privacy testing suites. *Privacy and Identity Management for Life*, pages 32–44, 2012.

[8] I. Agrafiotis, S. Creese, and M. Goldsmith. Formalising requirements for a biobank case study using a logic for consent and revocation. *Privacy and Identity Management for Life*, pages 232–244, 2012.

[9] I. Agrafiotis, S. Creese, M. Goldsmith, and N. Moffat. Requirements formalisation. Technical report, Technical Report, University of Oxford, Department of Computer Science, 2012.

[10] I. Agrafiotis, S. Creese, M. Goldsmith, and N. Papanikolaou. Reaching for informed revocation: Shutting off the tap on personal data. *Privacy and Identity Management for Life*, pages 246–258, 2010.

[11] I. Agrafiotis, S. Creese, M. Goldsmith, and N. Papanikolaou. Applying formal methods to detect and resolve ambiguities in privacy requirements. *Privacy and Identity Management for Life*, pages 271–282, 2011.

[12] R. Agrawal, P. Bird, T. Grandison, J. Kiernan, S. Logan, and W. Rjaibi. Extending relational database systems to automatically enforce privacy policies. In *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, pages 1013–1022. IEEE, 2005.

[13] B.J. Alge, G.A. Ballinger, S. Tangirala, and J.L. Oakley. Information privacy in organizations: Empowering creative and extrarole performance. *Journal of Applied Psychology*, 91(1):221, 2006.

[14] A.L. Allen. *Uneasy access: Privacy for women in a free society*. Rowman & Littlefield Pub Inc, 1988.

[15] A.L. Allen. Coercing privacy. *Wm. & Mary L. Rev.*, 40:723, 1998.

[16] I.O. Angell and S. Smithson. *Information Systems Management*. Macmillan, London, 1991.

[17] G.J. Annas and M.A. Grodin. *The Nazi doctors and the Nuremberg Code: human rights in human experimentation*. Oxford University Press, USA, 1995.

[18] C. Ardagna, E. Damiani, S. De Capitani di Vimercati, and P. Samarati. Towards privacy-enhanced authorization policies and languages. *Data and Applications Security XIX*, pages 924–924, 2005.

[19] C.A. Ardagna, M. Cremonini, S. De Capitani di Vimercati, and P. Samarati. A privacy-aware access control system. *Journal of Computer Security*, 16(4):369–397, 2008.

[20] C.A. ARDAGNA, S. DE CAPITANI DI VIMERCATI, E. Pedrini, and P. Samarati. Primelife policy language. In *W3C workshop on access control application scenarios: 17-18 november 2009, Luxemburg: proceedings*. Dipartimento di Ingegneria dell'informazione e metodi matematici, Università degli studi di Bergamo; W3C, 2009.

[21] A. Armando, R. Carbone, L. Compagna, K. Li, and G. Pellegrino. Model-checking driven security testing of web-based applications. In *Software Testing, Verification, and Validation Workshops (ICSTW), 2010 Third International Conference on*, pages 361–370. IEEE, 2010.

[22] P. Ashley, S. Hada, G. Karjoth, C. Powers, and M. Schunter. Enterprise privacy authorization language (epal). *Research report*, 3485, 2003.

[23] N.F. Awad and MS Krishnan. The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS quarterly*, pages 13–28, 2006.

[24] P.D. Barbara Downe-Wamboldt RN. Content analysis: method, applications, and issues. *Health care for women international*, 13(3):313–321, 1992.

[25] J. Barrigar, J. Burkell, and I. Kerr. Let's not get psyched out of privacy: Reflections on withdrawing consent to the collection, use and disclosure of personal information. *Can. Bus. LJ*, 44:54, 2006.

[26] A. Barth, A. Datta, J.C. Mitchell, and H. Nissenbaum. Privacy and contextual integrity: Framework and applications. In *Security and Privacy, 2006 IEEE Symposium on*, pages 15–pp. IEEE, 2006.

[27] E. Beardsley. Privacy: Autonomy and selective disclosure. *Nomos XIII: Privacy*, pages 56–70, 1971.

[28] T.L. Beauchamp and J.F. Childress. *Principles of biomedical ethics*. Oxford University Press, USA, 2001.

[29] M. Becker, A. Malkis, and L. Bussard. A practical generic privacy language. *Information Systems Security*, pages 125–139, 2011.

[30] M.Y. Becker. Secpal formalization and extensions. Technical report, Citeseer, 2009.

[31] M.Y. Becker, C. Fournet, and A.D. Gordon. Secpal: Design and semantics of a decentralized authorization language. *Journal of Computer Security*, 18(4):619–665, 2010.

[32] M.Y. Becker, A. Malkis, and L. Bussard. A framework for privacy preferences and data-handling policies. *Microsoft Research Cambridge Technical Report, MSR-TR-2009-128*, 2009.

[33] F. Bélanger and R.E. Crossler. Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, 35(4):1017–1041, 2011.

[34] D.E. Bell and L.J. La Padula. Secure computer system: Unified exposition and multics interpretation. Technical report, DTIC Document, 1976.

[35] P. Benassi. Truste: an online privacy seal program. *Communications of the ACM*, 42(2):56–59, 1999.

[36] S.I. Benn. Privacy, freedom, and respect for persons. *Nomos XIII: Privacy*, 1:26, 1971.

[37] C.J. Bennett. Cookies, web bugs, webcams and cue cats: Patterns of surveillance on the world wide web. *Ethics and Information Technology*, 3(3):195–208, 2001.

[38] E. Bertino. Rbac modelsconcepts and trends. *Computers & Security*, 22(6):511–514, 2003.

[39] P. Beynon-Davies. The uk national identity card. *Journal of Information Technology Teaching Cases*, 1(1):12–21, 2011.

[40] E.J. Bloustein. Privacy as an aspect of human dignity: An answer to dean prosser. *NYUL Rev.*, 39:962, 1964.

[41] E.J. Bloustein. Group privacy: The right to huddle. *Rutgers-Cam LJ*, 8:219, 1976.

[42] S. Bok. *Secrets: On the ethics of concealment and revelation*. Vintage, 1989.

[43] P.A. Bonatti, E. Damiani, S. De Capitani di Vemercati, and P. Samarati. A component-based architecture for secure data publication. In *Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual*, pages 309–318. IEEE, 2001.

[44] J.A. Boydston. John dewey: The middle works, 1899-1924. *Vol*, 9:129, 1980.

[45] L.D. Brandeis and S.D. Warren. The right to privacy. *Harv. L. Rev.*, 4:193, 1890.

[46] M. Brown and R. Muchira. Investigating the relationship between internet privacy concerns and online purchase behavior. *Journal of Electronic Commerce Research*, 5(1):62–70, 2004.

[47] L. Bussard, G. Neven, and F.S. Preiss. Downstream usage control. In *Policies for Distributed Systems and Networks (POLICY), 2010 IEEE International Symposium on*, pages 22–29. IEEE, 2010.

[48] J.W. Byun and N. Li. Purpose based access control for privacy protection in relational database systems. *The VLDB Journal*, 17(4):603–619, 2008.

[49] B.J. Calder. Focus groups and the nature of qualitative marketing research. *Journal of Marketing Research*, pages 353–364, 1977.

[50] https://consult.justice.gov.uk/digital-communications/data-protection-proposals-cfe, 2012.

[51] L.J. Camp. Design for trust. *TRUST, REPUTATION AND SECURITY: THEORIES AND PRACTICE, Rino Falcone, ed., Springer-Verlang (Berlin), 2003*, 2003.

[52] Marco Casassa Mont, Siani Pearson, Gina Kounga, Yun Shen, and Pete Bramhall. Privacy and identity management in europe.overview of existing assurance methods in the area of privacy and it security. Technical report, HP Labs, Bristol, 2004.

[53] S. Castano, M.G. Fugini, G. Martella, and P. Samarati. Database security. *ACM Press Books, Wokingham, England: Addison-Wesley,— c1995*, 1, 1995.

[54] F.H. Cate. *Privacy in the information age*. Brookings Inst Pr, 1997.

[55] C.U. Ciborra and G.F. Lanzara. Formative contexts and information technology: Understanding the dynamics of innovation in organizations. *Accounting, management and information technologies*, 4(2):61–86, 1994.

[56] V. Ciriani, S. De Capitani Di Vimercati, S. Foresti, G. Livraga, and P. Samarati. Enforcing confidentiality and data visibility constraints: an obdd approach. *Data and Applications Security and Privacy XXV*, pages 44–59, 2011.

[57] V. Ciriani, S.D.C. Vimercati, S. Foresti, and P. Samarati. k-anonymous data mining: A survey. *Privacy-preserving data mining*, pages 105–136, 2008.

[58] E. Clarke and X. Zhao. Analyticaa theorem prover in mathematica. *Automated DeductionCADE-11*, pages 761–765, 1992.

[59] E.M. Clarke and R.P. Kurshan. Computer-aided verification. *Spectrum, IEEE*, 33(6):61–67, 1996.

[60] E.M. Clarke and J.M. Wing. Formal methods: State of the art and future directions. *ACM Computing Surveys (CSUR)*, 28(4):626–643, 1996.

[61] R. Clarke. Introduction to dataveillance and information privacy, and definitions of terms. *Roger Clarke's Dataveillance and Information Privacy Pages*, 1999.

[62] Federal Trade Commission. Children's online privacy protection act. http://www.ftc.gov/ogc/coppa1.htm, 1998.

[63] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle. The platform for privacy preferences 1.0 (p3p1. 0) specification. *W3C recommendation*, 16, 2002.

[64] L.F. Cranor. *Web privacy with P3P*. O'Reilly Media, 2002.

[65] L.F. Cranor, J. Reagle, and M.S. Ackerman. *Beyond concern: Understanding net users' attitudes about online privacy*. Cambridge, MA: MIT Press, 2000.

[66] J.B. Craven. Personhood: The right to be let alone. *Duke Law Journal*, 1976(4):699–720, 1976.

[67] M.J. Culnan. How did they get my name?: An exploratory investigation of consumer attitudes toward secondary information use. *Mis Quarterly*, 17(3):341–363, 1993.

[68] M.J. Culnan and P.K. Armstrong. Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, pages 104–115, 1999.

[69] M.J. Culnan and C.C. Williams. How ethics can enhance organizational privacy: Lessons from the choicepoint and tjx data breaches. *Mis Quarterly*, 33(4):673–687, 2009.

[70] L. Curren. Data protection law and the legal basis of privacy. *EnCoRe Project Briefing Paper*, 2009.

[71] L. Curren and J. Kaye. Revoking consent: A [] blind spot'in data protection law? *Computer Law & Security Review*, 26(3):273–283, 2010.

[72] A. Davies. Invading the mind: The right to privacy and the definition of terrorism in canada. *University of Ottawa Law & Technology Journal*, 3(1), 2006.

[73] S. Davies. Private virtue. *The Guardian*, 2002.

[74] Judith DeCew. Privacy. http://plato.stanford.edu/archives/fall2008/entries/privacy/, 2008.

[75] J.W. DeCew. *In pursuit of privacy: Law, ethics, and the rise of technology*. Cornell Univ Pr, 1997.

[76] I. Dey. *Qualitative data analysis: A user-friendly guide for social scientists*. Routledge, 1993.

[77] T. Dinev and P. Hart. An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1):61–80, 2006.

[78] http://www.theregister.co.uk/2009/12/21/dna_pnc, 2010.

[79] T. Dowty and D. Korff. Protecting the virtual child the law and childrens consent to sharing personal data. *The Nuffield Foundation*, 2009.

[80] EC. Commission proposes a comprehensive reform of the data protection rules. http://ec.europa.eu/justice/newsroom/dataprotection/news/120125_en.htm, 2012.

[81] S. Elo and H. Kyngäs. The qualitative content analysis process. *Journal of advanced nursing*, 62(1):107–115, 2008.

[82] T.I. Emerson. *The system of freedom of expression*, volume 3. Random House New York, 1970.

[83] http://www.encore-project.info.

[84] http://www.encore-project.info/newsletters/newsletter01/EnCoReJuly2010.htm.

[85] R.A. Epstein. Legal regulation of genetic discrimination: Old responses to new technology, the. *BUL Rev.*, 74:1, 1994.

[86] A. Etzioni. *The limits of privacy*. Basic Books (AZ), 1999.

[87] A. Etzioni. Are new technologies the enemy of privacy? *Knowledge, Technology & Policy*, 20(2):115–119, 2007.

[88] http://www.telegraph.co.uk/technology/internet/9223930/EU-cookie-law-will-cost-businesses-10billion.html, 2012.

[89] R.R. Faden, T.L. Beauchamp, and N.M.P. King. *A history and theory of informed consent*. Oxford University Press, USA, 1986.

[90] J. Feinberg. *Harm to self*, volume 3. Oxford University Press, USA, 1989.

[91] E.F. Fern. The use of focus groups for idea generation: the effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research*, pages 1–13, 1982.

[92] J.C. Fernandez, C. Jard, T. Jéron, and C. Viho. An experiment in automatic generation of test suites for protocols with verification technology* 1. *Science of Computer Programming*, 29(1-2):123–146, 1997.

[93] S. Fischer-Hubner. *IT-security and privacy: design and use of privacy-enhancing security mechanisms*. Springer-Verlag, 2001.

[94] B. Fisse and J. Braithwaite. *The impact of publicity on corporate offenders*. State Univ of New York Pr, 1983.

[95] Organisation for Economic Co-Operation and Development. Guidelines: On the protection of privacy and transborder of personal data. http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html, 1980.

[96] http://www.cabinetoffice.gov.uk/news/information_commissioner_conference_francis_maude_keynote_speech, 2012.

[97] C. Fried. *An anatomy of values: problems of personal and social choice*. Harvard University Press, 1970.

[98] B. Friedman, E. Felten, and L.I. Millett. Informed consent online: A conceptual model and design principles. *University of Washington Computer Science & Engineering Technical Report 00–12–2*, 2000.

[99] B. Friedman, P.H. Khan Jr, and D.C. Howe. Trust online. *Communications of the ACM*, 43(12):34–40, 2000.

[100] R. Garrett. The nature of privacy. *Philosophy Today*, 18(19741):264–84, 1974.

[101] R. Gavison. Privacy and the limits of law. *Yale LJ*, 89:421, 1979.

[102] J. Genova. *Wittgenstein: A way of seeing*. Psychology Press, 1995.

[103] A. Gibbs. Focus groups. *Social research update*, 19(8), 1997.

[104] B.G. Glaser and A.L. Strauss. *The discovery of grounded theory: Strategies for qualitative research*. Aldine de Gruyter, 1967.

[105] Blakley B. Glazer I. Privacy, identity and privacy strategies in-depth research overview, 2009.

[106] S. Godik, T. Moses, A. Anderson, B. Parducci, C. Adams, D. Flinn, G. Brose, H. Lockhart, K. Beznosov, M. Kudo, et al. extensible access control markup language (xacml) version 1.0. *OASIS standard, February*, 2003.

[107] E.L. Godkin. Libel and its legal remedy. *J. Soc. Sci.*, 12:69–80, 1880.

[108] E.L. Godkin. The rights of the citizen-iv-to his own reputation. *Scribners Magazine*, 8(1):58–68, 1890.

[109] M. Goldsmith et al. Fdr: User manual and tutorial, version 2.77. *Formal Systems (Europe) Ltd*, 2001.

[110] M. Goldsmith and I. Zakiuddin. Critical systems validation and verification with csp and fdr. *Applied Formal MethodsFM-Trends 98*, pages 243–250, 1999.

[111] G. Goncalves and A. Poniszewska-Maranda. Role engineering: From design to evolution of security schemes. *Journal of Systems and Software*, 81(8):1306–1326, 2008.

[112] J. Grimmelmann. Facebook and the social dynamics of privacy. *Iowa Law Review*, 95(4):1–52, 2009.

[113] M. Grodin. Historical origins of the nuremberg code. *Medicine, Ethics and the Third Reich: Historical and Contemporary Issues*, pages 169–194, 1994.

[114] H. Gross. Concept of privacy, the. *NYUL Rev.*, 42:34, 1967.

[115] I.H. Hann, K.L. Hui, T.S. Lee, and I.P.L. Png. Online information privacy: Measuring the cost-benefit trade-off. In *23rd International Conference on Information Systems*, pages 1–8, 2002.

[116] M.A. Harrison, W.L. Ruzzo, and J.D. Ullman. Protection in operating systems. *Communications of the ACM*, 19(8):461–471, 1976.

[117] K. Heubusch and S. Dortch. Big brother on the internet. *American Demographics*, 19(2):22, 1997.

[118] C. Hine. Privacy in the marketplace. *The Information Society*, 14(4):253–262, 1998.

[119] R.F. Hixson. *Privacy in a public society: Human rights in conflict*. Oxford University Press New York, 1987.

[120] C.A.R. Hoare. An axiomatic basis for computer programming. *Communications of the ACM*, 12(10):576–580, 1969.

[121] C.A.R. Hoare. Communicating sequential processes. *Communications of the ACM*, 21(8):666–677, 1978.

[122] K. Hoeyer. Informed consent: The making of a ubiquitous rule in medical practice. *Organization*, 16(2):267–288, 2009.

[123] D.L. Hoffman, T.P. Novak, and M. Peralta. Building consumer trust online. *Communications of the ACM*, 42(4):80–85, 1999.

[124] G. Hosein. Challenges in privacy advocacy. *Reinventing Data Protection?*, pages 253–261, 2009.

[125] I. Hosein. Privacy as freedom. *Human rights in the global information society*, pages 121–148, 2006.

[126] H.F. Hsieh and S.E. Shannon. Three approaches to qualitative content analysis. *Qualitative health research*, 15(9):1277–1288, 2005.

[127] K. Hui, H.H. Teo, and S. Lee. The value of privacy assurance: An exploratory field experiment. *Management Information Systems Quarterly*, 31(1):19, 2007.

[128] S. Hunter and L. Curtice. Working with adults with incapacity. *Private and confidential?: handling personal information in the social and health services*, page 191, 2008.

[129] http://www.ico.gov.uk/for_organisations/privacy_and_electronic_communications/the_guide/cookies.aspx, 2012.

[130] British Computer Society Specialist Interest Group in Software Testing (BCS SIGIST). Standard for software component testing. Technical report, British Computer Society, Working Draft 3.4, 2001.

[131] J.C. Inness. *Privacy, intimacy, and isolation*. Oxford University Press, USA, 1996.

[132] L. Introna and A. Pouloudi. Privacy in the information age: Stakeholders, interests and values. *Journal of Business Ethics*, 22(1):27–38, 1999.

[133] L.D. Introna. Privacy and the computer: why we need privacy in the information society. *Metaphilosophy*, 28(3):259–275, 1997.

[134] Fisher J.A. Procedural misconceptions and informed consent: insights from empirical research on the clinical trials industry. *Kennedy Institute of Ethics Journal*, 16:251–268, 2006.

[135] E. Jackson. *Medical law: text, cases, and materials*, volume 604. Oxford University Press Oxford, 2006.

[136] C.B. Jones. *Systematic software development using VDM*, volume 2. Prentice Hall, 1990.

[137] S.M. Jourard. Some psychological aspects of privacy. *Law and Contemporary Problems*, 31(2):307–318, 1966.

[138] F. Kafka. *The Trial*. Penguin Modern Classics, 1925.

[139] D. Kahneman, J.L. Knetsch, and R.H. Thaler. Experimental tests of the endowment effect and the coase theorem. *Journal of political Economy*, pages 1325–1348, 1990.

[140] B. Kaplan and J. Maxwell. Qualitative research methods for evaluating computer information systems. *Evaluating the Organizational Impact of Healthcare Information Systems*, pages 30–55, 2005.

[141] G. Karjoth, M. Schunter, and M. Waidner. Platform for enterprise privacy practices: Privacy-enabled management of customer data. In *Privacy Enhancing Technologies*, pages 194–198. Springer, 2003.

[142] K.L. Karst. Privacy and freedom. *Cornell L. Rev.*, 53:744, 1967.

[143] J. Katz. The nuremberg code and the nuremberg trial. *JAMA: the journal of the American Medical Association*, 276(20):1662–1666, 1996.

[144] P.G. Kelley, L. Cesca, J. Bresee, and L.F. Cranor. Standardizing privacy notices: An online study of the nutrition label approach. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1573–1582. ACM, 2010.

[145] T. Kelly and R. Weaver. The goal structuring notation–a safety argument notation. In *Proc. DSN 2004 Workshop on Assurance Cases*. Citeseer, 2004.

[146] I. Kerr. If left to their own devices... how drm and anti-circumvention laws can be used to hack privacy. *IN THE PUBLIC INTEREST: THE FUTURE OF CANADIAN COPYRIGHT LAW, Michael Geist, ed., Irwin Law, 2005*, 2005.

[147] I. KERR. Lessons from the identity trail: anonymity, privacy and identity in a networked society (harback). *Recherche*, 67:02, 2009.

[148] J. Kitzinger. The methodology of focus groups: the importance of interaction between research participants. *Sociology of health & illness*, 16(1):103–121, 1994.

[149] J. Kitzinger. Qualitative research: introducing focus groups. *Bmj*, 311(7000):299–302, 1995.

[150] J. Knodel. The design and analysis of focus group studies: A practical approach. *Successful focus groups: Advancing the state of the art*, 1:35–50, 1993.

[151] A. Kobsa. Tailoring privacy to users needs 1. *User Modeling 2001*, pages 301–313, 2001.

[152] J. Kolter, R. Schillinger, and G. Pernul. A privacy-enhanced attribute-based access control system. *Data and Applications Security XXI*, pages 129–143, 2007.

[153] N.L. Kondracki, N.S. Wellman, and D.R. Amundson. Content analysis: Review of methods and their applications in nutrition education. *Journal of Nutrition Education and Behavior*, 34(4):224–230, 2002.

[154] K. Krippendorff. *Content analysis: An introduction to its methodology*. Sage Publications, Inc, 2004.

[155] R.A. Krueger and M.A. Casey. *Focus groups: A practical guide for applied research*. Sage, 2009.

[156] RP Kurshan. *Computer-aided verification of coordinating processes: the automata-theoretic approach*. Princeton Univ Pr, 1994.

[157] J. Lach. the new gatekeepers. *American Demographics*, 21(6):41–42, 1999.

[158] K. LeFevre, R. Agrawal, V. Ercegovac, R. Ramakrishnan, Y. Xu, and D. DeWitt. Limiting disclosure in hippocratic databases. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pages 108–119. VLDB Endowment, 2004.

[159] D. Lin and M.C. Loui. *Taking the byte out of cookies: privacy, consent, and the Web*, volume 28. ACM, 1998.

[160] C. Liu, J.T. Marchewka, J. Lu, and C.S. Yu. Beyond concern: a privacy–trust–behavioral intention model of electronic commerce. *Information & Management*, 42(1):127–142, 2004.

[161] J. Locke. *The second treatise of government: and, A letter concerning toleration*. Dover Pubns, 2002.

[162] G. Lowenstein and J. Elster. *Choice over time*. Russell Sage Foundation, 1992.

[163] N. Lundblad and B. Masiello. Opt-in dystopias. *SCRIPTed*, 7(1):155–165, 2010.

[164] X. Luo. Trust production and privacy concerns on the internet: A framework based on relationship marketing and social exchange theory. *Industrial Marketing Management*, 31(2):111–118, 2002.

[165] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.

[166] C.A. MacKinnon. *Toward a feminist theory of the state*. Harvard Univ Pr, 1989.

[167] N.K. Malhotra, S.S. Kim, and J. Agarwal. Internet users' information privacy concerns(iuipc): the construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, 2004.

[168] G.T. Marx. Soft surveillance: The growth of mandatory volunteerism in collecting personal information" hey buddy can you spare a dna?". *Surveillance and security: Technological politics and power in everyday life*, pages 37–56, 2006.

[169] J. Mason. *Qualitative researching*. Sage Publications Ltd, 2002.

[170] J.A. Maxwell. *Qualitative research design: An interactive approach*. Sage Publications, Inc, 2005.

[171] V. Mayer-Schonberger. *Delete: The Virtue of Forgetting in the Digital Age (New in Paper)*. Princeton Univ Pr, 2011.

[172] P. Mayring. Qualitative content analysis. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, volume 1, 2000.

[173] HJ McCloskey. Privacy and the right to privacy. *Philosophy*, 55(211):17–38, 1980.

[174] L. McCreary. What was privacy? *Harvard business review*, 86(10):123, 2008.

[175] S. McRobb and S. Rogerson. Are they really listening?: An investigation into published online privacy policies at the beginning of the third millennium. *Information Technology & People*, 17(4):442–461, 2004.

[176] A.R. Miller. *The assault on privacy: Computers, data banks, and dossiers*. University of Michigan Press, 1971.

[177] L.I. Millett, B. Friedman, and E. Felten. Cookies and web browser design: Toward realizing informed consent online. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 46–52. ACM, 2001.

[178] G.R. Milne and M.J. Culnan. Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3):15–29, 2004.

[179] G.R. Milne and M.E. Gordon. Direct mail privacy-efficiency trade-offs within an implied social contract framework. *Journal of Public Policy & Marketing*, pages 206–215, 1993.

[180] R. Milner. *A calculus of communicating systems*. Springer-Verlag New York, Inc., 1982.

[181] A.D. Miyazaki and A. Fernandez. Consumer perceptions of privacy and security risks for online shopping. *Journal of Consumer Affairs*, 35(1):27–44, 2001.

[182] M. Moniruzzaman, M.S. Ferdous, and R. Hossain. A study of privacy policy enforcement in access control models. In *Computer and Information Technology (ICCIT), 2010 13th International Conference on*, pages 352–357. IEEE, 2010.

[183] M.C. Mont, S. Pearson, G. Kounga, Y. Shen, and P. Bramhall. On the management of consent and revocation in enterprises: Setting the context. Technical report, Technical Report HPL-2009-49, HP Labs, Bristol, 2009.

[184] M.C. Mont, V. Sharma, S. Pearson, R. Saeed, and M. Filz. Technical architecture arising from the third case study. Technical report, HP Labs, Bristol, 2011.

[185] D.L. Morgan. *Successful focus groups: Advancing the state of the art*, volume 156. Sage Publications, Inc, 1993.

[186] D.L. Morgan. Focus groups. *Annual review of sociology*, pages 129–152, 1996.

[187] T. Moses et al. extensible access control markup language (xacml) version 2.0. *Oasis Standard*, 200502, 2005.

[188] R.S. Murphy. Property rights in personal information: An economic defense of privacy. *Geo. LJ*, 84:2381, 1995.

[189] M.D. Myers and DE Avison. Qualitative research in information systems. *Management Information Systems Quarterly*, 21:241–242, 1997.

[190] K.A. Neuendorf. *The content analysis guidebook*. Sage Publications, Inc, 2002.

[191] Q. Ni, E. Bertino, and J. Lobo. An obligation model bridging access control policies and privacy policies. In *Proceedings of the 13th ACM symposium on Access control models and technologies*, pages 133–142. ACM, 2008.

[192] Q. Ni, E. Bertino, J. Lobo, and S.B. Calo. Privacy-aware role-based access control. *Security & Privacy, IEEE*, 7(4):35–43, 2009.

[193] H. Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

[194] U.S. Department of Health and Human Services. Summary of the hipaa privacy rule. http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/index.html, 1996.

[195] P. Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA L. Rev.*, 57:1701, 2010.

[196] W. Orlikowski and J.J. Baroudi. Studying information technology in organizations: Research approaches and assumptions. *IOMS: Information Systems Working Papers*, 1990.

[197] M. Ortendahl and J.F. Fries. Time-related issues with application to health gains and losses. *Journal of clinical epidemiology*, 55(9):843–848, 2002.

[198] G. Orwell. *1984*. Secker and Warburg, 1949.

[199] http://wyvern.ndcls.ox.ac.uk/orb/.

[200] D. ONeil. Analysis of internet users level of online privacy concerns. *Social Science Computer Review*, 19(1):17–31, 2001.

[201] N. Papanikolaou, S. Creese, M. Goldsmith, M.C. Mont, and S. Pearson. Encore: Towards a holistic approach to privacy. In *Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on*, pages 1–6. IEEE, 2010.

[202] W.A. Parent. Privacy, morality, and the law. *Philosophy & Public Affairs*, 12(4):269–288, 1983.

[203] J. Park and R. Sandhu. The ucon abc usage control model. *ACM Transactions on Information and System Security (TISSEC)*, 7(1):128–174, 2004.

[204] J. Park and R. Sandhu. A position paper: A usage control (ucon) model for social networks privacy. *w3.org*, 2010.

[205] R.B. Parker. Definition of privacy, a. *Rutgers L. Rev.*, 27:275, 1973.

[206] A. Paschke. Eca-lp/eca-ruleml: A homogeneous event-condition-action logic programming language. *Arxiv preprint cs/0609143*, 2006.

[207] E. Perecman and S.R. Curran. *A handbook for social science field research: essays & bibliographic sources on research design and methods*. Sage Publications, Inc, 2006.

[208] A. Pnueli. The temporal logic of programs. In *Foundations of Computer Science, 1977., 18th Annual Symposium on*, pages 46–57. IEEE, 1977.

[209] R.A. Posner. *Economic analysis of law*. Aspen Publishers, 1973.

[210] R.A. Posner. Right of privacy, the. *Ga. L. Rev.*, 12:393, 1977.

[211] R.A. Posner. *The economics of justice*. Harvard Univ Pr, 1983.

[212] R.A. Posner. *Not a suicide pact: The Constitution in a time of national emergency.* Oxford University Press, USA, 2006.

[213] R.C. Post. Three concepts of privacy. *Geo. LJ*, 89:2087, 2000.

[214] S. Preibusch. Experiments and formal methods for privacy research. *Privacy and Usability Methods Pow-wow (PUMP)*, 2010.

[215] W. Prosser. The torts of privacy. *California Law Review*, 383(48):392–98, 1960.

[216] Data protection working party. Article 29: Opinion 15/2011 on the definition of consent. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp187_en.pdf, 2011.

[217] J. Queille and J. Sifakis. Specification and verification of concurrent systems in cesar. In *International Symposium on Programming*, pages 337–351. Springer, 1982.

[218] J. Rachels. Why privacy is important. *Philosophy & Public Affairs*, 4(4):323–333, 1975.

[219] TB Riley. Security vs. privacy: A comparative analysis of canada, the united kingdom and the united states. *Journal of Business and Public Policy*, 1(2):1–21, 2007.

[220] A.W. Roscoe. Model-checking csp. *A Classical Mind, Essays in Honour of CAR Hoare. Prentice-Hall*, pages 353–378, 1994.

[221] D.J. Rothman. *Strangers at the bedside: A history of how law and bioethics transformed medical decision making.* Aldine de Gruyter, 2003.

[222] J. Rubenfeld. The right of privacy. *Harvard Law Review*, pages 737–807, 1989.

[223] J.B. Rule. *Privacy in peril.* Oxford University Press, USA, 2007.

[224] R. Sandhu and P. Samarati. Authentication, access control, and audit. *ACM Computing Surveys (CSUR)*, 28(1):241–243, 1996.

[225] R.S. Sandhu, E.J. Coyne, H.L. Feinstein, and C.E. Youman. Role-based access control models. *Computer*, 29(2):38–47, 1996.

[226] R.S. Sandhu and P. Samarati. Access control: principle and practice. *Communications Magazine, IEEE*, 32(9):40–48, 1994.

[227] F.D. Schoeman. *Philosophical dimensions of privacy: An anthology.* Cambridge Univ Pr, 1984.

[228] P.M. Schwartz. Privacy and democracy in cyberspace. *Vand. L. Rev.*, 52:1607, 1999.

[229] http://www.telegraph.co.uk/news/newstopics/politics/lawandorder/5898488/Sex_offender_register_for_life_breaches_rights_of_rapists_and_paedophiles.html, 2010.

[230] K.B. Sheehan. Toward a typology of internet users and online privacy concerns. *The Information Society*, 18(1):21–32, 2002.

[231] H.A. Simon. *Models of bounded rationality: Empirically grounded economic reason*, volume 3. The MIT Press, 1997.

[232] H.J. Smith, T. Dinev, and H. Xu. Information privacy research: An interdisciplinary review. *MIS Quarterly*, 35(4):989–1015, 2011.

[233] H.J. Smith, S.J. Milberg, and S.J. Burke. Information privacy: measuring individuals' concerns about organizational practices. *MIS quarterly*, pages 167–196, 1996.

[234] D. Solove. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3):477, 2006.

[235] D. Solove. 'ive got nothing to hide' and other misunderstandings of privacy. *San Diego Law Review*, 44:745–748, 2007.

[236] D. Solove. Understanding privacy. *Daniel J. Solove, UNDERSTANDING PRIVACY, Harvard University Press, May 2008, GWU Legal Studies Research Paper No. 420, GWU Law School Public Law Research Paper No. 420*, 2008.

[237] D.J. Solove. *The digital person: Technology and privacy in the information age*, volume 1. NYU Press, 2004.

[238] D.J. Solove. *The future of reputation: Gossip, rumor, and privacy on the Internet*. Yale Univ Pr, 2007.

[239] J.Y. Son and S.S. Kim. Internet users information privacy-protective responses: A taxonomy and a nomological model. *Mis Quarterly*, 32(3):503–529, 2008.

[240] J.M. Spivey. *Understanding Z: a specification language and its formal semantics*, volume 3. Cambridge Univ Pr, 1988.

[241] D.W. Stewart, P.N. Shamdasani, and D.W. Rook. *Focus groups: Theory and practice*, volume 20. Sage Publications, Inc, 2007.

[242] K.A. Stewart and A.H. Segars. An empirical examination of the concern for information privacy instrument. *Information Systems Research*, 13(1):36–49, 2002.

[243] W.J. Stuntz. Secret service: Against privacy and transparency. *New Republic*, 12:12, 2006.

[244] L. Sweeney et al. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty Fuzziness and Knowledge Based Systems*, 10(5):557–570, 2002.

[245] C.J. Sykes. The end of privacy: Personal rights in the surveillance society. *The end of Privacy: Personal Rights in the Surveillance Society*, 1999.

[246] Y.H. Tan and W. Thoen. Toward a generic model of trust for electronic commerce. *International Journal of Electronic Commerce*, 5(2):61–74, 2000.

[247] http://maude.cs.uiuc.edu/.

[248] S. Trabelsi, A. Njeh, L. Bussard, and G. Neven. The ppl engine: A symmetric architecture for privacy policy handling. In *W3C Workshop on Privacy and data usage control*, volume 4, 2010.

[249] M.C. Tremblay, A.R. Hevner, and D.J. Berndt. Focus groups for artifact refinement and evaluation in design research. *Communications of the Association for Information Systems*, 26(1):27, 2010.

[250] J.Y. Tsai, S. Egelman, L. Cranor, and A. Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2):254, 2011.

[251] M. Tschantz and J. Wing. Formal methods for privacy. *FM 2009: Formal Methods*, pages 1–15, 2009.

[252] Identity Assurance (IdA) Programme Statements UK Cabinet Office. http://services.parliament.uk/hansard/Commons/ByDate/20110518/writtenministerialstatements/part003.html, 2011.

[253] http://www.theregister.co.uk/2008/08/20/uk_gov_lost_records/ Lastaccessedon06/02/2010, 2010.

[254] E. Van Den Haag. On privacy. *Privacy: Nomos XIII. New York: Atherton*, 1971.

[255] T.P. Van Dyke, V. Midha, and H. Nemati. The effect of consumer privacy empowerment on trust and privacy concerns in e-commerce. *Electronic Markets*, 17(1):68–81, 2007.

[256] W3C. A p3p preference exchange language 1.0, 2002.

[257] W3C. P3p 1.1 specification, 2006.

[258] J. Waldo, H. Lin, and L.I. Millett. *Engaging privacy and information technology in a digital age*. Natl Academy Pr, 2007.

[259] K.K. Waterman. Pre-processing legal text: Policy parsing and isomorphic intermediate rep-resentation. In *Intelligent Information Privacy Management Symposium at the AAAI Spring Symposium*, 2010.

[260] R.P. Weber. *Basic content analysis*, volume 49. Sage Pubns, 1990.

[261] B. Week. A little net privacy, please. *Business Week*, 1998.

[262] A.F. Westin. Privacy and freedom. *Washington and Lee Law Review*, 25(1):166, 1968.

[263] http://www.whitehouse.gov/sites/default/files/email-files/privacy_white_paper.pdf, 2012.

[264] E.A. Whitley. The identity project: an assessment of the uk identity cards bill and its implications, 2005.

[265] E.A. Whitley. Information privacy consent and the 'control' of personal data. *Inform. Secur. Tech. Rep.*, DOI:10.1016/j.istr.2009.10.001, 2009.

[266] E.A. Whitley. Perceptions of government technology, surveillance and privacy: the UK identity cards scheme. *New directions in surveillance and privacy*, page 133, 2009.

[267] E.A. Whitley and N. Kanellopoulou. Privacy and informed consent in online interactions: Evidence from expert focus groups, 2010.

[268] M. Williams. *Making sense of social research*. Sage Publications Ltd, 2003.

[269] TM Williamson. Research, informed consent and the limits of disclosure. *Bioethics*, 15(4):341–363, 2001.

[270] P. Willis and M. Trondman. Manifesto for ethnography. *Ethnography*, 1(1):5–16, 2000.

[271] L. Wittgenstein. Philosophical investigations, trans. *GEM Anscombe*, 261:49, 1953.

[272] R.K. Yin. *Case study research: Design and methods*, volume 5. Sage publications, INC, 2009.

[273] X. Zhang, F. Parisi-Presicce, R. Sandhu, and J. Park. Formal model and policy specification of usage control. *ACM Transactions on Information and System Security (TISSEC)*, 8(4):351–387, 2005.

[274] D.L. Zimmerman. Requiem for a heavyweight: A farewell to warren and brandeis's privacy tort. *Cornell L. Rev.*, 68:291, 1982.