Grant Agreement No.: 611001

# UCN


Instrument: Collaborative Project
Call Identifier: FP7-ICT-2013-10
Objective 1.6: Connected and Social Media

# D 4.1: Requirements, Ethics and Security Models for Privacy Preserving Data Management

Due date of deliverable: 30.09.2014

Actual submission date: 30.09.2014

Start date of project: October 1st 2013 Duration: 36 months

Project Manager: Henrik Lundgren, Technicolor SA

Author(s): EURECOM, NICTA, University of Nottingham

Editor: University of Nottingham

Revision: v.1.0

## Abstract

This deliverable uses a variety of resources to explore user understandings of privacy with regard to their online activity and the range of user expectations and requirements and ethical considerations UCN will need to take into account in order to design a platform that properly addresses user privacy concerns. It then presents the way in which the UCN environment will be constituted in order to meet people's privacy requirements and the kinds of security models UCN will adopt in order to ensure privacy is preserved at all points.

| Project co-funded by the European Commission in the 7th Framework Programme (2007-2013) | | |
|---|---|---|
| Dissemination Level | | |
| PU | **Public** | ✓ |
| PP | **Restricted to other programme participants (including the Commission Services)** | |
| RE | **Restricted to a group specified by the consortium (including the Commission Services)** | |
| CO | **Confidential, only for members of the consortium (including the Commission Services)** | |

# Table of Contents

# 1 INTRODUCTION

In this deliverable, findings from both previously reported research and more recently conducted ethnographic studies are used in order to explore user understandings of privacy when they are engaged in various kinds of online activity. These materials form the basis of a discussion of user expectations and requirements regarding matters of privacy and ethics that UCN will need to take into account when designing a system architecture for user-centric connected media services. The deliverable then articulates the ways in which the UCN environment will be developed in order to meet the various privacy requirements we have outlined and the kinds of security models that are going to be adopted in order to maximise the scope for preserving user privacy across a range of different kinds of circumstances.

## *1.1 BACKGROUND*

The basic underlying idea of UCN is to exploit user-generated data to facilitate much greater personalization of online digital services, with a particular emphasis upon media. It will do this by designing improved *content recommendation* and *content delivery frameworks*. One of the issues the project is aiming to address is the current tendency for information pertaining to recommendation to be bound up with single service providers, even where multiple providers may be providing a similar service, making it impossible to spread recommendation across the whole provision of that service. The other main issue to address is the fact that users are continually moving across a whole range of different contexts and ecologies of consumption such that the whats, wheres, whens, hows and whys of what gets consumed are subject to a shifting pattern of preferences and concerns that are poorly serviced by existing delivery mechanisms.

An inevitable feature of these principal ambitions is that much more personal information about users has to be acquired than might previously have been the case. To pin down people's content preferences and interests involves taking a much closer look at the things people currently consume, the contexts in which they consume them, and the patterns of interaction that surround such consumption, including their interactions via social media. To properly examine the ways in which people might like best to consume content involves taking a close look at their routines and habits, the ways they like doing things, the things they like to use for doing those things, etc.

Set against this are certain expectations regarding privacy and security. Despite the fact that there has been a huge increase in the kinds of data that can now be accumulated regarding people and their interactions with technology, many privacy debates continue to revolve around fairly traditional concerns such as: the need to provide mechanisms for maintaining privacy when conducting digital transactions (e.g. Beach et al, 2009; Bellotti & Sellen, 1993; Boyle & Greenberg, 2005; Chakraborty et al, 2013; Clarke et al, 2012; Dwyer et al, 2007; Ghani & Sidek, 2008 & 2009; Kapadia et al, 2007; Raij et al, 2007; Schrammel et al, 2009); the blurring of boundaries between what counts as private or public (e.g. Barnes, 2006; Gandy Jr, 1993; Oka et al, 2011); the extent to which incentives to reveal personal information may serve to undermine people's privacy (e.g. Livingstone, 2008; Sleeper et al, 2013; Wang et al, 2011); and the need to ensure that appropriate measures are put in place to adequately inform

people of their rights and to obtain their consent (e.g. Boyle & Greenberg, 2005; Conti et al, 2013; Milberg et al, 1995; Tang et al, 2011).

In the case of security, the focus is on ensuring that private or personal information is not revealed and that, when it is revealed, it is only revealed to those with an express right to access it (e.g. Clarke et al, 2012; Song et al, 2013). This is in turn is often related to concerns with potential security breaches such as the accidental revelation of personal information or its deliberate misuse.

Because UCN needs to access and use personal information it is necessary for it to ensure that the consortium properly understands the kinds of concerns about privacy and security people are likely to make manifest and to have in place mechanisms for addressing those concerns in an appropriate fashion.

## *1.2 OBJECTIVES*

In view of the above observations this deliverable is geared towards several key objectives:

- Mapping out the landscape of user understandings about privacy and security against which its research and design objectives will be situated.
- Elaborating from these understandings the requirements for effective privacy and security management the design elements of the project will need to address.
- Indicating the ways in which the project will be seeking to address these requirements (split here across three main security models: where there is no access to user data at all; where there access to only some user data; and where there is more or less full access to user data).

## *1.3 APPROACH*

In order to address the user understanding elements of the above objectives we shall be making use of two principal resources:

- A very extensive pre-existing literature on people's concerns about privacy and their relationship to the use of information technology.
- A collection of recent ethnographic studies that have explored, amongst other things, how people orient to privacy and make their concerns about it manifest in the course of their ordinary everyday activities.

The literature we shall be examining comes primarily, though by no means exclusively, from the field of Human-Computer Interaction (HCI). It is inevitably at the point where human beings and machines rub up against one another that issues like privacy will come to the fore so this is the domain where privacy has been most extensively discussed and examined. It will be seen that a variety of perspectives are found within this literature. Some are primarily focused upon technical matters and design; some are intensely sociological and psychological in character, seeking to understand privacy in terms of human phenomena; some are centred upon economic considerations and how to address design and human issues relating to privacy in terms of good business practice; and some are principally interested in how privacy might be tackled as a matter of policy and good organisational practice.

The ethnographic materials come from three sets of studies in domestic environments. One of these bodies of data was gathered in the context of trying to understand people's use of wireless networks in their homes and the ways in which that use was embedded in everyday practice and the social organisation of household environments (see Crabtree et al, 2012 & 2014). This data was derived from studies across 24 extremely diverse households including families with younger and older children, older couples, single and shared occupancy households, living in many different kinds of housing, covering a variety of income brackets and with a range of different occupational backgrounds. Equally diverse data from a further 16 households is being used from a second set of studies where the principal objective was to map people's online activity in terms of their consumption and use of digital services (see Tolmie et al, 2013). Finally, we are making use of data gathered in an as yet unreported set of studies across 4 different households where a variety of sensors were installed with the express aim of collecting personal data in order to probe how people would incorporate the collection of such data into their everyday lives and how they would orient to the sharing of that data with other parties (see HAT, 2014).

The privacy requirement and security model aspects of the work presented here are founded upon an extensive body of existing research and, as will be seen in section 3, are formulated principally around the notion of Personal Information Hubs (or PIHs).

Right from the outset UCN emphasized the importance of confidentiality to users:

> In any privacy preserving data management system, the very first privacy requirement for the user is data confidentiality. A user initially encrypts the data it wishes to outsource and uploads it to the (distributed) system. (Original Project Proposal)

However, the problems with traditional responses to this were also identified:

> While classical encryption algorithms are considered a good candidate for this requirement, they often obstruct the second advantage of a powerful data management system which is computation outsourcing. Indeed, once its data is uploaded into the system, an authorized node (such as a recommender system) may want to query the data for either a lookup or a retrieval of some information while still assuring privacy. The node may of course download the whole encrypted data, decrypt it and perform the required operation but such a naïve solution unfortunately does not take advantage of the computational capabilities of the overall data management system. (Original Project Proposal)

The potential relevance here of Private Information Retrieval (PIR) was discussed:

> Current solutions based on either Private information retrieval PIR) and/or searchable encryption have received a lot of attention. In PIR (private information retrieval) (Cachin et al, 1999; Chor et al, 1995; Goldreich & Ostrovsky, 1996; Ostrovsky & Skeith, 2007; Sion & Carbunat, 2007), a user retrieves a specific data from a database located in a server. The only privacy goal in PIR is access privacy whereby the server should not discover which data a user is interested in. Note that PIR does not ensure privacy of data in the database. Furthermore,

> searchable encryption (Boneh et al 2004; Boneh et al, 2007; Chang & Mitzenmacher, 2005; Curtmola et al, 2006; Ogata & Curosawa, 2004; Song et al, 2000) allows a node to verify whether some specific "keywords" exist in the remote data. With such techniques, user privacy is guaranteed thanks to the encryption of the queries and the stored data. (Original Project Proposal)

However, a need to go beyond PIR in UCN was also identified:

> Most of the previously mentioned PIR and searchable encryption techniques target single server settings. Therefore, existing PIR and searchable encryption techniques will be revisited in order to design dedicated primitives for a distributed environment. Furthermore, the only "privacy" goal in PIR is access privacy whereby the server should not discover which data a user is interested in. Therefore PIR alone is not sufficient to assure storage privacy. Similarly, in existing searchable encryption solutions the result (which is a Boolean result) originating from a search query is known to the adversary. Hence, the node that processes the query can obtain some knowledge of a user's queries. In a dedicated security model, the adversary should sometimes not even learn anything about queries or results. Finally, the majority of privacy preserving lookup schemes do not consider the data management system as being malicious while executing the required processing operation. The new privacy preserving primitives proposed in UCN will give the user the ability to verify the correctness of operations. (Original Project Proposal)

Another area of potential concern that was identified was the preservation of privacy across aggregated data. Here, too, some relevant approaches have already been located:

> Secure data aggregation has been studied in the context of wireless sensor networks (Onen & Molva, 2007; Castellucia et al, 2009) whereby nodes aggregate their data to the current intermediate aggregate they receive in order to forward the resulting value to the next hop towards the sink. Thanks to the use of homomorphic encryption algorithms, nodes can perform correct aggregation operations over encrypted data. Homomorphic encryption (ElGamal, 1985; Paillier, 1999; Boneh et al, 2005) allows a third party to perform meaningful computation over encrypted data. Most of the very well-known homomorphic encryption techniques support very limited operations such as simple addition {Paillier, 1999) or multiplication (ElGamal, 1985) or both (Boneh et al, 2005). (Original Project Proposal)

Once again in UCN there is a need to extend beyond this:

> Current secure aggregation techniques rely on the existence of a trusted central node (the sink) and assume the correct behavior of all participating nodes. In UCN, aggregation operations have to be distributed to all nodes among which some of them may be malicious. Therefore techniques such as secure multi-party computation (SMC) (Cramer et al, 2001; Yao, 1982; Ben-Or et al, 1988; Bendlin et al, 2011) will be investigated in order to assure the correctness of the resulting aggregate even with the existence of some misbehaving nodes. Finally, the newly proposed privacy preserving and verifiable aggregation mechanisms will

> consider the trade-off between security and efficiency by considering more realistic adversary models. (Original Project Proposal)

A further relevant approach that UCN has been investigating is *differential privacy*. Differential Privacy is a technique developed by Dwork et al. at Microsoft Research (Chawla et al, 2005; Dwork, 2006a; Dwork, 2006b; Dwork, 2008a; Dwork, 2008b), which bounds information rate leakage from queries over data. An earlier similar system was IBM Research's Hippocratic Database, for storing patient-privilege medical records. The advantage with differential privacy is that it:

> … lets one build a set of tools for managing responses to statistical queries over data which must still be secured perhaps with mandatory role based access control privileges, in such a way that the user cannot reconstruct more precise identification of principles in the dataset beyond a formally well-specified bound. This means that personally sensitive data such as Internet packet traces or social network measurements can be shared between researchers without invading personal privacy, and that assurances can be made with accuracy. The rate of queries must be limited, and the range from upper to lower bound of any value to be protected known, a priori. The limit on the number of roles and the number of queries per role must also be known ahead of time. Nor does the system necessarily stop inference given other external data about subjects in the database. The assumption is generally made that the user of the data is "honest, but curious" as opposed to being a genuine adversary with disclosure of identity and associated attributes in mind. In other words, a user such as an advertiser or market researcher who wishes to target adverts, get click through statistics, and find out generally about the preferences of users over some property (perhaps location), can be provided. As an added incentive for the data user to behave correctly, one can log accesses to a third party, so that misbehaviour can be publicly shown (and therefore potentially negatively impact the overly intrusive advertisers' or excessively invasive market research analytics' business). In some cases, simple fuzzing of data may suffice. (Original Project Proposal)

The aim in UCN will be to move beyond current state-of-the-art approaches and develop a system specifically tailored to the privacy requirements of providing tailored services and effective recommendation to people across a wide variety of different settings where a foundational need to preserve core aspects of privacy will remain paramount.

## *1.4 DOCUMENT STRUCTURE*

As indicated above, the aim of this document is to outline user requirements regarding privacy and to present the basic premises upon which UCN will seek to address these requirements within the design of its system. In order to do this we will:

- Examine user understandings of privacy, both in terms of the ordinary everyday assumptions people make about privacy and in terms of the extent to which current privacy mechanisms are intelligible to them when they are engaged in online activity.
- Make explicit user requirements and expectations regarding privacy, both in terms of what data is being collected about them and in terms of how that data might be used.

- Present how these various considerations will be addressed through the privacy and security models being developed within UCN and how these will be open to flexible tailoring to meet the requirements of different kinds of context.

In the next section we shall use a range of existing work and related ethnographic studies in order to discuss just how people currently understand and orient to privacy as an everyday feature of their use of computing technology.

# 2 USER UNDERSTANDINGS OF PRIVACY

This section of the deliverable is devoted to outlining the findings from existing bodies of research regarding how users of computing systems typically understand and orient to privacy as a feature of using those systems. In order to do this it looks at three principal themes: first of all how people make use of their ordinary everyday understandings of privacy and security in the context of online interactions; and secondly, the ways in which privacy and security mechanisms already in place in computer systems are intelligible to users or otherwise. These materials will form the basis of our later discussions of requirements and how UCN will approach the creation of effective privacy and security mechanisms that will address them.

## *2.1 ORDINARY UNDERSTANDINGS OF PRIVACY*

An important part of the background to the work being undertaken in UCN is the extent to which people are already encountering a range of technologies that monitor and measure their locations and activities in various ways. The use of Web 2.0 for a variety of purposes and the massive uptake of social media such as Facebook, Twitter, and YouTube have already begun to frame the privacy debates here. This has been further elaborated by the roll-out of clamp-based meters such as the Current Cost device and the growing popularity of devices such as the Fitbit, the Withings 'Smart-Body Analyzer' and 'Pulse', and smartphone apps such as http://sleepyti.me which have led to burgeoning discussion about lifelogging technologies and the 'quantified self'. As efforts proceed towards the realisation of the Internet of Things the scope for monitoring even more of what we do across the environments we inhabit adds further fuel to concerns about what all of this might mean for our privacy.

A deeper concern here is the extent to which people are aware of the fact their data might be used in the first place (Acquisti & Gross, 2006; Balebako et al, 2013; Conti et al, 2013; Ellison et al, 2007; Hawkey & Inkpen, 2006; Song et al, 2013), something visible in a number of the recent debates about spying and hacking. This also points to a broader issue that we shall be looking at here regarding how people may generally understand and orient towards matters relating to privacy (Friedman, 1997; Lin et al, 2012). More than this, in view of the fact that a number of studies point to how reasoning changes according to the specific situations where information is disclosed or hidden, it is also important to look at people's *situated* understanding to get a proper sense of how they really reason about privacy (Froehlich et al, 2007; Sleeper et al, 2013; Viswanath et al, 2008; Wang et al, 2011).

### *2.1.1 THE RISK OF EXPOSURE*

Concerns about privacy and risks of exposure when online have been around for some time (e.g. see Ackerman et al, 1999) but the increasing range of opportunities for data gathering in our environments has instilled new vigour into this debate.

In this section we shall be exploring the matter of privacy from a variety of different angles in order to map out the landscape within which discussions of this topic have been situated. First of all we will look at ways in which privacy has been framed as a general background concern in relation to the use of technology and computing systems. We will then look at how some of the perceived risks associated with online activity have been characterised. After that we will review a range of distinctions that have been drawn regarding how people orient to privacy

according to the different kinds of devices they use. This is followed by an examination of the various trade-off behaviours visible in how people will disclose personal information in order to accomplish various ends. This in turn leads into a discussion of how notions of personal or sensitive information encompass a wide variety of different kinds of data that may be oriented to in quite distinct ways. A further refinement of the variable sensitivity of information is the fact that just what counts as sensitive or not is subject to a variety of contingent and situated considerations, rendering the generic classification of information in terms of privacy a problematic enterprise.

#### *2.1.1.1 A MATTER OF GENERAL CONCERN*

Some of the commentary to be encountered regarding privacy is hugely general and never really gets into the meat of what the perceived disadvantage of exposure might be, insisting instead upon a presumed 'wrongness' of other people getting hold of information about you without your knowledge, your permission or your capacity to exercise control. This broad moral and ethical position often then forms the backdrop for other kinds of debate.

Olson et al (2005) note that the tendency for technological change to provoke debate about possible 'violations of privacy' is actually far from being a recent phenomenon. Indeed, an impetus to develop proper privacy legislation can be traced back to at least the 1800s when the rapid evolution of photographic techniques led to concerns about pictures being taken of people without their permission (Warren & Brandeis, 1890). The spread of telephones throughout domestic settings prompted a whole new round of debate in the 1920s (Fischer, 1992). So the scope for information about people to be collected through the development of ubiquitous computing systems and networked sensing technology needs to be situated against this historical background of a general unease about what technology might serve to reveal about people and how it might be thus abused. Broad findings from surveys that people are concerned about online privacy (Ackerman et al, 1999; Cranor et al, 1999) or that they consider it to be an 'important issue' (GVU, 1998; TRUSTe, 2011; Ur et al, 2005) should therefore be considered in these terms.

Whilst not quite going into exactly what people fear might happen as a consequence of the exposure of personal information, some authors do break down the general concern into various sub-categories. Smith et al (1996), for instance, propose the following: 'concerrns about collection of personal information'; concerns about 'processing errors'; concerns about the 'further use of personal data (control)'; and concerns about 'improper access to the information'. Yao et al (2007) echo these worries about 'unauthorized secondary use', 'access', 'collection' and 'errors'. Metzger and Docter (2003) propose a somewhat different (though related) set of four categories: 'anonymity'; 'intrusion'; 'surveillance'; and 'autonomy'. Boyle and Greenberg (2005), centring their interest upon the question of control, suggest yet another set of related (and partially overlapping) concerns: 'solitude' (e.g. the right to no longer be visible or monitored in any way); 'confidentiality' (specifically 'control over the fidelity with which others access information about you'); and 'autonomy' (e.g. 'control over one's own behavior and the expression of identity'). This matter of 'control' features in a number of other discussions of privacy as well (e.g. Ackerman et al, 1999; Fox et al, 2000).

In a related set of discussions one can also encounter investigations of the different 'types' of people who may have different kinds of concerns about privacy and the different kinds of

demographics to which these people may belong. One of the classic typologies in this vein was first coined by Westin (1998). Here a strong distinction is drawn between people who are 'privacy fundementalists', 'privacy unconcerned' or 'privacy pragmatists'. This typology has subsequently been incorporated within the analysis of several other teams investigating privacy (e.g. Ackerman et al, 1999; Berendt et al, 2005; Hawkey & Inkpen, 2006; Olson et al, 2005; P&AB, 2003; Taylor, 2003). Here the focus is upon distinguishing between people who are generally unwilling to share data about themselves, people who have little or no concern about doing so, and people who engage in some kind of assessment of the relative pros and cons of sharing information and proceed accordingly.

Some analysts have drilled into this typology a little further. Berendt et al (2005), for instance, examining disclosure to an artificial agent, add in a further distinction for privacy pragmatists regarding "*identity concerned* users" who "are more concerned about revealing information like their name, email, or mailing address" and "*profiling averse* users" who "are more concerned about disclosing information such as their interests, hobbies, and health status." Hawkey and Inkpen (2006) suggest that the Westin-Harris segmentation model describes people's 'inherent privacy concerns' and claim that the "individual will have a large effect on his privacy comfort level in a given situation" and could even be used as a 'predictor' of people's privacy preferences. The variable situations they have in mind relate to matters such as "visible content, level of control, and viewers". As with other studies where an emphasis has been put on control they particularly focus on the amount of control people have over the information made available to others, notably at the point of physically entering information at a machine, e.g. "A high amount of control (e.g. control over input devices) should lessen privacy concerns, while lower levels of control should increase concerns… When participants envisioned themselves in control of the keyboard and mouse, they have the least amount of concern across the viewing audience. As control is lost, the amount of concern grows."

Very much aligned with the Westin-Harris model, but couched in terms of people's 'beliefs in privacy rights', others such as Yao et al (2007) claim (somewhat obviously) that "the more that people believe in the right to privacy and the more they desire privacy in the physical world, the more they are likely to have online privacy concerns (about both companies and other entities)." Yao et al, however, also introduce other determining features such as: psychological dispositions (such as a 'need for privacy' and 'self efficacy'); the amount people use the Internet; and people's 'fluency' with using computer systems. In a similar assessment of people's privacy concerns according to specific kinds of scenario, Culnan and Armstrong (1999) examined how people said they would behave according to whether they were told 'fair information practices' would be used or not when disclosing information for the purposes of online purchasing of products. Here the suggestion is that people who are concerned about privacy rights are more likely to withhold information when they do not have a guarantee of controlled use of their data, but will disclose the same amount of information as anyone else when such guarantees *are* forthcoming.

Another way in which researchers have sought to classify people with regard to their privacy concerns is according to various kinds of demographic considerations. Westin (1998), for instance, found that women were somewhat more concerned about online privacy than men, and this has been suggested across a number of other studies as well (e.g. Furash, 1997; Kehoe, Pitkow, & Morton, 1997; Milne & Rohm, 2000; O'Neil, 2001; and Sheehan, 1999). Yao et al (2007) argue that the results of all these studies are 'inconclusive' and claim on the

contrary, that "Gender has no direct or indirect impact on concerns about online privacy." Others focus their attention upon age demographics instead. Several studies here (e.g. Yao et al, 2007) ground their findings in an older body of research in developmental psychology regarding how people's attitudes to privacy change as they mature, with children over 10 becoming increasingly concerned about having their own private space (Lawton & Bader, 1970; Marhsall, 1974) and the need for 'privacy markers' and 'privacy rules' such as having locks on doors, placing signs on doors, or having people knock before they enter (Parke & Swain, 1979; Wolfe & Laufer, 1974). The suggestion here is that a concern with online privacy is something that evolves in line with these other privacy considerations. The implicit counterpoint to this argument is that children under a certain age have a reduced sensitivity to the possible risks of exposure. This way of thinking is embodied in another set of broad concerns regarding the *protection* of children when engaged in online activity. Several studies here have indicated that people are generally 'less comfortable' with children disclosing information than they would be themselves when disclosing identical kinds of data (contact details, age, etc.) (Ackerman et al, 1999; Cranor et al, 1999). These studies, however, do not engage in any assessment of how these kinds of concerns are bound up as much with other considerations such as the vulnerability of children to abuse as they are with 'privacy' per se.

A related argument here revolves around the capacity of children to exercise what one might term 'sound moral judgment' and the overall *morality* of certain kinds of data being exposed to others. Here the concern reflects a suggestion that certain kinds of disclosure may be deemed by most people to be inappropriate in some way (Boyle & Greenberg, 2005; Emanuel et al, 2013; Milberg et al, 1995; Raij et al, 2011; Schrammel et al, 2009; Sleeper et al, 2013; Wang et al, 2011). Discussions of the behaviour of adolescents using social media are particularly pertinent here, with a number of studies focusing especially upon issues such as inappropriate forms of self-display (Lenhart & Madden, 2007; Zhao et al, 2008), excessive or naïve revelations (Kwan & Skoric, 2013), and narcissistic behaviour (Livingstone, 2008).

A few studies have investigated concerns about online privacy from the point of view of cultural differences (e.g. Olson et al, 2005) and have indicated that there may be some specific issues here (for instance with regard to images and contact details).

Yet another body of studies has looked at the relationship between attitudes to privacy and people's experience of using the Internet. A number of these studies have suggested that the more frequently people use online resources, the less concerned they appear to be about privacy (e.g. Metzger, 2004; Phelps et al, 2000; UCLA Center for Communication Policy, 2000, 2001, 2003, 2004). However, the results of these studies are challenged by Yao et al (2007) who claim "the empirical link between Internet use fluency and beliefs in privacy rights is not clear."

New features of the privacy debate revolve around the range of information that may now be captured and the persistence of information such that it is no longer necessarily the case that things will simply disappear or be forgotten.

One of the discussions relevant to this debate is the notion of it being possible for data from a wide range of resources to be assembled together in what has been termed a person's *personal digital footprint* (Mortier et al, 2010). The concern in this regard relates to the possibility of an assemblage of what might individually amount to just small traces of people's activity in

the digital record in such a way that the totality of information can begin to become informative to others in unforeseen ways (Gemmell et al, 2004; Kapadia et al, 2007; Karkkainen et al, 2010; Sellen & Whittaker, 2010). All sorts of things may feed into the personal digital footprint including email, photos and videos, documents, system and activity logs, network logs, application logs, application preferences, cache and cookie information, personal profiles, social media content, GPS data, GPS trails, geotags, physical recordings, sensor data, transaction data, health records, government records, financial databases, and so on.

Reinforcing the above considerations, Spiekermann (2005) showed participants in a study a video of potemtial applications of RFID technology and found that people were concerned about things such as the 'further use' of the collected data, a 'perceived helplessness' regarding the capacity of such data to be gathered and used, and the ease with which such data might be used. She noted in particular that "participants were concerned over a loss of control over the technology and uncertainties regarding the technology's utility and effective operation". Much of the current angst about these issues is encapsulated in the following:

> "Consumers and social media users deal with data collection that is invisible. Card swiping and form signing are now replaced with practically nothing (Lahlou et al, 2005). Data collection done online is the equivalent of watching a person walk around a store and keep track of every object they look at. Many of these websites may talk together to form a user profile. Sites such as Amazon, Google, and Facebook bring together information on an unprecedented scale. Users have little say in the matter. In order to use these services, you consent to this type of information sharing. This type of profiling already exists. For example, Gmail reads email messages to recommend products and services to be posted in advertisements. Soon, Google plans to change the way it uses data to integrate its different products (web searches, Gmail, Google+) so that information can be gathered across the different applications for use in advertisement. Such a change is unprecedented and currently being challenged by several US States (Vijayen, 2012)."
>
> (Blasbalg et al, 2012)

In a broader analysis Adams and Sasse (1999) suggest that there are "four factors that determine the perception of privacy in richly sensed environments." These are: the recipient of the data; the context in which the data is collected; the potential sensitivity of what may be sensed; and the kinds of use that might be made of the data.

A further notable vein of interest with regard to recent technological change and its impact on privacy relates to mobile location-based applications. Holone and Herstad (2010), for instance, building upon previous work by Palen and Dourish (2003), argue that a key issue is the persistence of information collected by such applications. They suggest that people currently work on the assumption that most of the time information about where they are is of an essentially ephemeral character and that the capacity of such applications to preserve indefinitely traces of where they have been creates completely new tensions between privacy and the scope for such information to be shared with others. This is something we shall look at again in section 2.1.1.3.

#### *2.1.1.2 SPECIFIC THREATS PERCEIVED*

Much of the concern about privacy one encounters in various investigations of its nature is framed around perceived risks of either having information disclosed to the wrong people or in damaging ways, or else around risks of losing control of personal information in some way so that its use is no longer available to you.

Whilst, as noted above, a lot of the research about loss of control is quite general in nature, some research does drill a little more deeply into exactly what kinds of risks people think may confront them if they should lose control of their data in some way. The specific risks people may associate with having their personal information exposed to third parties are diverse. One such concern is with how exposure may result in damage to people's identities in some way (Emanuel et al, 2013; Lenhart & Madden, 2007; Livingstone, 2008; Zhao et al, 2008), such as identity theft, where a malicious party impersonates a victim (Palen & Dourish, 2003). Other concerns may be seen to attach to the risk of exposing what might be termed 'negative information' or conveying a negative impression, with this in turn being used for various dubious ends such as shame and exposure, retribution, bullying, or even blackmail (Greenberg & Rounding, 2001; Kobsa et al, 2012; Kwan & Skoric, 2013; Palen & Dourish, 2003; Patil et al, 2012; Sleeper et al, 2013). Palen and Dourish (2003) also suggest that people have concerns about surveillance from a variety of sources (including employers) and unwanted government inspection of their activity. Others such as Klasnja et al (2009) point to concerns people have regarding the risk of financial loss as a result of conducting online transactions and note that a number of people adopt certain strategies such as not undertaking online purchases or online banking in public locations, to try and limit the risk of having their accounts hacked.

A particular focus in the HCI literature has centred upon the risks online activity may pose for effective impression management. Nissenbaum (2004), for instance, suggests that people manage the disclosure of private information according to certain assumed 'norms' regarding what is appropriate within any interaction and the capacity of information to travel. She subsumes these norms within a general concept of 'contextual integrity' and claims that it is when contextual integrity is breached that people feel their privacy has been violated in some way. Several authors (e.g. Wang et al, 2011; Patil et al, 2012) use this as a basis for analysing situations of failed impression management in the context of people's interactions on social media, especially where information intended for a very restricted audience is accidentally exposed to a much wider circle of acquaintance. Patil et al (2012) argue that this is a much more serious risk than many other kinds of risk associated with online activity. Situating this within the specific risks of unwanted location disclosures they make the following observations:

> "… the situations of regrettable disclosure for many types of private information (e.g., financial information, passwords, etc.) involve commercial entities and/or unknown third parties, such as spammers, phishers, and hackers. In such cases, the consequences are limited to the individual whose information was disclosed and future damage can often be prevented (via actions like closing accounts, changing passwords, etc.). In the case of unintended location disclosures, however, the consequences often result from contextual aspects and lead to social repercussions that can affect multiple parties and often linger on in the future."
>
> (Patil et al, 2012)

Examples cited of the kinds of negative consequences that can accrue to unwanted disclosures include:

> "I was out with some friends. I told my other friend, who wanted me to hang out with her, that I wasn't feeling good (sick in bed) because I did not want to go to the place she was going. I checked into another bar that night forgetting what I had told my friend, who is on Facebook, that I wasn't going out at all."
> "My boss saw where I was when I told her I was sick and I got fired."
> "It made my girlfriend jealous because I checked into a local restaurant with my female co-worker."
> "My wife saw that I was at the mall buying a gift when I stated I was somewhere else. It ruined the surprise."
>
> (Patil et al, 2012)

Watson et al (2012) and a range of others (Karr-Wisniewski et al, 2011; Lapinen et al, 2009; Stutzman & Harzog, 2012; Wisniewski et al, 2012) note that, as consequence of this perceived risk, a large number of people have become extremely careful about what they post on social network sites.

In a somewhat different assessment of the potential risks arising from disclosure on social networking sites, Mao et al (2011) look at the risks associated with revealing more than is intended via posts on Twitter. Here they identify things like burglars "automatically receiving alerts about vacation messages", "law enforcement … receiving alerts about drunk driving", and "insurance agencies … receiving alerts about people with medical conditions".

#### *2.1.1.3 DIFFERENCES ACROSS DEVICES AND APPLICATIONS*

Another kind of differentiation one can see in studies of privacy relates to how the use of different kinds of devices may impact upon how people reason about privacy, for instance when using smartphones as opposed to desktop computers or laptops (Chin et al, 2012), though there has been little research to date on how this may also relate to the use of tablet computers. Related discussions here can also be bound up with things like the use of different kinds of networks, or even specific kinds of device-bound activity such as text-messaging. Building upon the above sections, a further set of associated discussions relate to the use of specific kinds of applications such as social media, web browsing or location-based services. We will look at each of these different strands of research separately.

##### *2.1.1.3.1 Devices*

One strand of research has focused on the orientations of **laptop** users to matters of privacy. In particular several researchers (e.g. Hawkey & Inkpen, 2006) have noted that laptop users express more concerns about privacy with regard to those specific devices than about desktop PCs. This is put down to a greater degree of personal browsing taking place on laptops than on workplace PCs and the added risk of laptops being used in many different locations. Other accounts suggest that, when it comes to portable devices, laptops are more likely to be used for sensitive tasks such as financial transactions than phones (Chin et al, 2012).

An especially large body of device-focused research on privacy has been devoted to the use of **mobile phones** and, in particular smartphones and smartphone applications. Much of this research suggests that users have notable concerns about exposure on their phones and some, such as Chin et al (2012), argue that users avoid using certain applications on their

smartphones precisely because they consider the risk of exposure too high. A range of features has been found to prompt concern, including applications drawing upon people's contact lists, the use of people's locations (Urban et al, 2012), and requests for personal information prior to a number of applications even being installed (Boyles et al, 2012). Indeed, several researchers report a significant number of people (up to 57%) uninstalling applications from their smartphones the moment they encounter these kinds of requests (Balebako et al, 2013; Boyles et al, 2012). Other researchers also report significant numbers of people (up to 60%) avoiding financial transactions on their phones because of a perceived personal security risk (Chin et al, 2012). Other reported concerns include the use of Social Security Numbers, accessing bank accounts, and accessing health or medical records. By contrast people seem to exhibit few privacy concerns regarding the use of financial management tools, health and fitness management tools, the sharing of photos, or accessing email (Chin et al, op cit). Elaboration of just what the exact nature of the risks of using phones might be includes matters such as the invisible storage of personal information in applications on the phone, the relative instability of phone applications, the possibility of phones being hacked in public locations, and the greater scope for losing phones and other people finding them. Further concerns also relate to a distrust of public WiFi and 3G networks. It should be stressed, however, that reports also suggest that potentially sensitive activities are not undertaken on phones simply because people have more trouble with the interface and do not want to risk making errors when engaged in tasks they consider to be important (Chin et al, op cit).

As a separate matter researchers have also reported people being especially inclined to describe their phones as 'private' or 'personal' devices (e.g. Hakkila & Chatfield, 2005). Some put this down to the range of personal information etc that people routinely store on their phones (e.g. pictures, text messages, phone calls, emails, locations, phone numbers, calendars, etc.) (Ben-Asher et al, 2011; Chin et al, 2012; Hakkila & Chatfield, 2005). Studies here also indicate people having a similar orientation to their phones and other people's phones as they do to other personal artefacts such as wallets or purses, with a reluctance to even touch or answer other people's mobile phones without their express permission (Hakkila & Chatfield, op cit). Hakkila & Chatfield (op cit) also find that people are generally reluctant to give other people permission to to answer their phone. An interesting refinement of the reasoning they mention here is that people would make an exception if it were a family member answering a call from another family member. This exhibition of contingent reasoning is something we shall return to in section 2.1.1.6.

#### *2.1.1.3.2 Channels*

Another slightly distinct set of concerns have been previously uncovered relating not so much to the specific device being used as to the *channel*. Research here is somewhat older but is still indicative of a refinement in reasoning that has been reported elsewhere (e.g. Tolmie, 2010). Westin (1991), for instance, uncovered a definite distinction regarding the degree to which people saw marketing and advertising as an invasion of privacy between postal mail and telephone calls, with the latter being seen as much more problematic. Cranor et al (1999) also found a distinction between how much personal information people were prepared to give to receive investment advice via websites or postal mail, with the latter being seen to be much less problematic. Thus there is some evidence to suggest that reasoning, at least with regard to the appropriateness of contact, does also encompass the particular choice of channel.

#### *2.1.1.3.3 Applications*

Like the more general use of mobile phones, another area that has received a fair amount of attention in the context of privacy research is the use of **instant messaging and text messaging**. Several studies have looked at both of these applications alongside one another, with many findings being found to relate to both of them equally (e.g. Grinter & Palen, 2002; Hakkila & Chatfield, 2005; Ito & Daisuke, 2003). Some specific features are singled out for their relevance to concerns about privacy such as there being no overtly visible trace (assuming deletion), the marking out of availability and unavailability to other people in your social group, and the opportunity to engage in private conversations without being overheard. Some aspects are also an extension of the orientation to mobile phones already discussed above, for instance not looking at text messages on other people's phones even when they are right next to you unless invited to do so, with answering somebody else's text message being almost completely taboo (Hakkila & Chatfield, 2005) (though empirical observations show that there are also exceptions to this, e.g. parents responding to texts from their children interchangeably on the same phone). Patil and Kobsa (2004) and Hakkila and Chatfield (2005) also found people in widely divergent settings equally adamant that they expected only the intended recipient of instant messaging and text messages to look at the messages they sent and to generally treat their messages as confidential, at the same time noting that this is entirely dependent upon the discretion of the recipient. In this respect the concerns here elide into the previous discussion regarding channel selection, though here it is a matter of medium, because a number of researchers have indicated that text messaging is generally considered a more 'private' form of communication than voice communication. Additional findings here are that some people do adopt added strategies to ensure the privacy of their text communications, e.g. deleting messages immediately after they have been read, using a different language, keeping hold of your phone at all times, using 'code' and slang, and so on. More people are reported, however, to assume that everyone understands the etiquette associated with texting and will therefore act in an appropriate fashion.

With the explosion in use of **social networking sites** there has been a concomitant interest in exploring the implications of using such sites for people's privacy. Much of this research has centred upon matters of impression management, as reported above. Other research, however, has focused on the limits of user understanding of what the real risks of disclosure might be when using such sites, especially with regard to organisational mining of such data (Blasbalg et al, 2012; Kisilevich & Mansmann, 2010; Vijayen, 2012). This touches again upon two recurrent themes in much of the research regarding privacy in the context of the increasing pervasiveness of data capturing technology: the lack of visibility of data collection (and therefore accountability for its use); and the extent to which users have at best a partial understanding of what technologies may be capable of. As will be made evident, both of these concerns are central to UCN's objective of developing a system that is properly attuned to people's requirements regarding privacy. Nor is UCN alone in this endeavour. Lehtinen et al (2009), for instance, noting that poor privacy safeguards limit the use of social networking sites by some cohorts within the population (especially older adults), point firmly to the need for much more transparent and straightforward privacy management tools in this context.

A further issue uncovered in recent research is the scope for even apparently public forums on social networking sites to become a source of accidental disclosure of sensitive information. Mao et al (2011) summarise matters thus:

> "Some may argue that by 'definition' information posted publicly on Twitter cannot be private and that Twitter users ought to realize that. A recent qualitative study of 'regrets' on Facebook shows that users "do not think about . . . the consequences of their posts" and they regret posts made when "they are in a 'hot' state of high emotion when posting, or under the influence of drugs or alcohol" (Wang et al, 2011). We … show there is a plethora of sensitive information revealed by Twitter users, not only about themselves but about other users. While users may themselves not think their posts are sensitive, we focus on categories of leaks where there is a clear potential for negative consequences to the user. In general it is hard to anticipate what other forms of leaks may occur from 'public' tweets, but recent news provides yet another note of caution. The New York Times reported on how various companies are now 'scoring' users based on Twitter (and other) feeds along various dimensions (NYT, 2011). Some applications of such scoring may be helpful for the user … but others may be especially harmful, when, for example, insurance companies score people based on their reported behaviors and increase premiums, or worse, deny them insurance."
>
> (Mao et al, 2011)

A somewhat distinct set of studies takes as its topic the ways in which people orient to privacy in the context of **browsing the web**. Returning to the above matter and the potential mining of data by large organizations, some studies (e.g. Panjwani et al, 2013) have suggested that a very large number of people (up to 84%) have search queries in their browser history that they deem potentially sensitive and that they are largely resitant to efforts by entities such as Google to track their search history for this reason, even if this places limits upon the scope for personalization. Ackerman et al (1999) looked at people's attitudes to long-term tracking of their online activity via cookies. Here the finding was that a significant number of people (52%) have concerns about cookies and manage their browser settings accordingly, but that there is also a significant number of people who do not have a good understanding of cookies or what the implications of accepting them might be.

Hawkey and Inkpen (2006) examined this issue from a different perspective, looking instead at people's sensitivities to the exposure of their web browsing activities to other people with whom they were acquainted to varying degrees, such as co-workers, bosses, or family members. Interestingly the suggestion is that in the context of their usual browsing habits many people are actually more uneasy about family members seeing their activity than co-workers, though this varied somewhat regarding the kind of content (medical searches might be more willingly exposed to family members, for instance, whilst erotica is deemed equally problematic for either of these cohorts). Hawkey and Inkpen also highlight a potential privacy issue relating to the increasingly large number of different devices people may use for accessing online resources:

> "People use a variety of computers regularly: laptops, single user PCs, and shared PCs, both at home and away from home. One problem with managing the privacy of traces of previous web browsing activity is that it is not always clear what traces will be revealed. With multiple devices, there may be increased uncertainty, particularly for those users that don't partition their browsing activities between locations and devices. Additionally, many participants indicated that they used their web browser convenience features differently for each computer. This lack of standardized settings across computers could add to the uncertainty about what will be revealed for each computer."
>
> (Hawkey & Inkpen, 2006)

Something we have already discussed to some extent in the context of both perceived threats and the use of smartphones is the increasingly large number of mobile applications that make **use of a user's location in some way**. Patil et al (2012) claim that an 'overwhelmingly' large number of people express a strong preference for only having their location made available upon their express instruction rather than as an automatic feature and this is echoed in a study by Chin et al (2012). However, they also acknowledge that this is becoming increasingly complicated as more and more device features turn upon the sharing of location in some way. Tsai et al (2010), for instance, list a variety of benefits people understand will accrue to their use of location-based sertvices, including: 'the safety of friends, coworkers, and children'; the coordination of activities and meetings; and 'finding people with similar interests'. We have already mentioned above some of the problems people anticipate may arise as a consequence of sharing their location. Other reported issues include: having to put up with targeted advertising; the exposure of addresses; 'being stalked'; and 'being tracked by the government of bosses' (Tsai et al, 2010). However, despite a number of studies that suggest location privacy is a significant issue for people using smartphones (see, for instance, Toch et al, 2010; Sadeh et al, 2009; and Ongtang et al, 2009), other studies have found that, despite there being a preference for more control, people are not unduly bothered if location-based services *are* activated (e.g. Chin et al, 2012).

A quite large literature within HCI is devoted to how people **share location information with one another** and with organisations and a number of studies examine how people manage privacy as a feature of this (Anthony et al, 2007; Barkhus, 2004; Barkhus & Dey, 2003; Consolvo et al, 2005; Cvrcek et al, 2006; Danezis et al, 2005; Iachello et al, 2005; Wiese et al, 2011). Consolvo et al (2005), for instance, suggest that disclosure of location information is very much tied up with how people anticipate the information will be used. Both Anthony et al (2007) and Consolvo et al (op cit) indicate that other considerations people bring to bear include where they actually are at the time, who it is who is asking, and the amount of detail about their current location they are likely to need to share. In relation to these kinds of findings, Holone and Herstad (2010) discuss the extent to which reasoning about location is embedded within existing social practices that do not trade upon the kind of detailed location tracking available in new technology and the extent to which location-based applications can therefore breach people's commonsense understanding. They particularly look at this from the point of view of assistive technologies where there is a trade-off to be considered between this and the potential for location tracking to improve care and support.

#### *2.1.1.3.4 Networks*

Another set of privacy-related issues that are not bound up with specific devices or applications, but which are still embedded within reasoning about the use of specific aspects of technology, is the matter of which network people are connected to. Both Chin et al (2012) and Hakkila and Chatfield (2005) find that this is in large part bound up with a general distrust of networks and the people who might be 'hanging around on them', rather than more specific considerations. Here is a sample of the kinds of things reported:

> "Any idiot with…$20 with malicious intent can pick up anything [over the air] from anyone with a cell phone."
> "I tend to use my phone in environments where hackers hang out. Just going to a techie cafe and giving away your credit card on your iPhone seems like asking for it."

> "If I'm on my phone, I'm probably out and I'm a little wary of using my credit card cuz there are plenty of hackers hanging out in cafes in SF."
>
> (Chin et al, 2012)

> "Anything on the internet, whether its email or blogging is not private. Anything that goes through a router isn't secure.'
>
> (Hakkila & Chatfield, 2005)

#### *2.1.1.4 A NECESSARY EVIL OR MAYBE EVEN AN OPPORTUNITY*

At the same time as one encounters a range of negative articulations about notions of privacy in the digital age, there are also new discussions developing regarding how the apparent thirst for data about people might be used by them to their own advantage by entering into various kinds of transactional engagements with the people who would seek to collect it. Indeed, a number of studies have shown that, whilst users may have a broad interest in managing the risk of exposure, there are also numerous rationales people will adopt for, on the contrary, quite explicitly making personal information available. In a previous deliverable (UCN D5.1, 2014) we have already noted that users are usually quite "pragmatic about the use of their data" and "aware that services that are of interest to them often require personal information and are willing to provide the data if they perceive a benefit and do not feel their data is used against their interest". Many of the considerations here are actually relatively mundane and more to do with the ordinary trade-offs one has to engage in in order to get by in life. So, and for instance, selected revelation of certain aspects of personal information is often necessary to proceed with a relationship, or is required to accomplish some other kind of end (Ackerman et al, 1999; Chakraborty et al, 2013; Chen & Xu, 2013; Chong & Treiblmaler, 2011). Examples here include gaining access to information or services (Joinson, 2008); obtaining enhanced services such as improved healthcare (Beach et al, 2009; Chen & Xu, 2013; Martin et al, 2013; Morris et al, 2011; Pratt et al, 2006; Raij et al, 2011; Taylor & Dajani, 2008); enhancing how you manage particular activities or pastimes (such as cooking and eating or sports activities) (Martin et al, 2013); or accessing features specifically tailored or personalised to meet your own preferences (Barua et al, 2011; Durrant et al, 2011; Krause et al, 2006; Sarin et al, 2008)[1]. In relation to all this Ackerman et al noted in 1999 an interest on the part of users to have an 'auto-fill' button that could be clicked on in browsers to enter recurrently requested information on web forms. Auto-fill is, of course, now a commonplace feature in most browsers and used pervasively. Much of this is bound up with a routine need to provide information to bring about transactions and obtain implicit value such as shopping benefits (Phelps et al, 2000) thus bringing about explicit rewards such as discounts, loyalty points or free gifts. It is also more pragmatically about the need to provide a billing address when purchasing from Amazon which you may choose to store to avoid having to re-enter it on subsequent occasions, along with traditionally more sensitive data such as credit card details (Sarin et al, 2008: Winckler et al, 2011).

Thus it can be seen that various modalities and situations exist under which users may well be perfectly willing to engage in the exchange or revelation of personal information, though usually with the expectation of some kind of resulting benefit (Blasberg et al, 2012), leading to some researchers terming these kinds of considerations a 'trading data for benefits privacy

[1] Though it is worth noting that surveys in the late 1990s (e.g. GVU, 1998) were finding that these kinds of exchange of personal data for benefit araangments were relatively unpopular, largely because of uncertainty about how the data might then be used.

calculus' (Dwyer et al, 2007). This kind of behaviour led Olson et al (2005) to comment that "typically, people do not want to keep everything private" and that they are "surprisingly willing to give away private information both for small amounts of money or privileges or even when talking with an anthropomorphic softbot on the Web (see also Hann et al, 2002, and Spiekermann et al, 2001). Some researchers are not only surprised by this but troubled by it as well. Berendt et al (2005), for instance, having been told by participants in their study that they were reluctant to disclose information online found that "the absolute level of disclosure was alarmingly high". This difference between stated preferences and behaviour had already been noted by Ackerman et al (1999), but other researchers, such as Blasberg et al (2012) and Chellappa (2002), explore the matter further and suggest that the prompts to this kind of diclosure are things like: a belief that the user will retain control over the information; that the requested information is apparently relevant to the activity being undertaken; and that it seems likely to the user that the provision of the information will lead to 'valid inferences' about their preferences. Blasberg et al (op cit), however, retain concerns about the exercise of these assumptions in certain domains:

> "In the case of social media, users feel they have control over their information as is evident by how much of it they voluntarily write on their postings. Many MySpace and Facebook users do not understand that if their profile is set to "private", their information can still be used and saved by the owners of the web site. This violates Chellappa's (2002) idea of perceived privacy. Chellappa suggests that the impression consumers have regarding the collection, access, use and disclosure of their private personal information is consistent with their beliefs regarding the way that information is being used."
>
> (Blasberg et al, 2012)

A related point we shall be returning to later on is the tension that exists between using personal information to enhance the ways systems work and the risk this may involve regarding possible exposure. Some of the research regarding attitudes to privacy suggests that there are certain aspects of computing support where users are fully aware of this kind of tension. Ur et al (2012), for instance, uncovered exactly this kind of articulation with regard to online behavioural advertising (OBA). However, Ur et al also found a number of ways in which people's understandings of how OBA works and the implications of that for their privacy were at odds with the actual technical operation of such systems, underscoring issues we raise in section 2.3 regarding the intelligibility of systems for users. Ur et al's work is based upon previous studies that had revealed a number of concerns about OBA (e.g. McDonald & Cranor, 2010; Purcell et al, 2012; Turow et al, 2009). However, other studies have found that concerns about OBA are significantly mitigated where users receive assurance that only non-personally identifiable information will be used (KPMG, 2011; Lendenmann, 2010), though certificating schemes for websites (such as the Advertising Option Icon in the US (TRUSTe, 2011)) seem to have marginal impact here (Ur et al, 2005).

Ackerman et al (1999) and Cranor et al (1999) also found that apparently fairly casual attitidues to sharing information online were not actually by any means wholesale and that there were certain situations where particular kinds of information would definitely be withheld. There was, for instance (echoing the above), a widespread reluctance to have personal information added to marketing lists and prospecting financial organisations would be examined for what they might have to offer but here, too, contact details would be withheld unless it was for the purposes of receiving specific information in the mail. They

found additionally that, whilst people would happily provide things like postcodes and answer questions about a wide range of interests and personal opinions for the purposes of online surveys, they exhibited much more reluctance to give more exact information such as their name. Attitudes to cookies were also found to be variable, with a willingness for them to be used to provide customization upon return to the same site, but a dislike of the possibility that they might be used to provide customized advertising.

Looking to information gathered by more ubiquitous computing type technologies, lifelogging data is also beginning to find similar kinds of trade-off uses, for instance in the area of P4 medicine – "Personalized, Predictive, Preventive, Participatory" (Hood, 2008; Hood & Friend, 2011; Sobradillo et al, 2011). Relatedly, Vendor Relationship Management (VRM) is a proposed use for centralised personal datastores where, inverting traditional notions of Customer Relationship Management, the user (customer) manages the data they expose to vendors such as their address, to control moments of change such as moving house http://cyber.law.harvard.edu/projectvrm/.

Other articulations are even more actively framed around notions of deliberately exploiting the exposure of data about oneself for positive advantage.

One such rationale might be the overall management of one's digital identity (Lenhart & Madden, 2007), often (following on from Goffman's (1959) notion of 'the presentation of self') for the purposes of positive impression management (Brody et al, 2012; Zhao et al, 2008) (we have already noted the problems associated with *failed* impression management above), whereby favourable impressions might be given to managers at work, prospective employers, etc. This extends to how people may seek to manage what is termed their 'super-identity' (their aggregate online identity across multiple sites) and the ways in which such identity management may lead to certain forms of empowerment.

More pragmatically, numerous uses bring benefits related to quite explicit commercial gain, whereby your personal data is ascribed a specific transaction value (Hoffman et al, 1999; Milberg et al, 1995; Mortier et al, 2010; Ng, 2013; Schwarz, 2003) and, when provided, will result in monetary reward such as a transfer to a PayPal account. Much of the transaction reasoning here is still structured around traditional notions of media copyright and Digital Asset Management (Austerberry, 2004), but there are increasingly varied discussions about how other kinds of personal information markets might be established (HAT, 2014).

#### *2.1.1.5 THE NATURE OF PERSONAL INFORMATION*

Some research on the topic of privacy devotes itself more to trying to differentiate between different categories of information and how people may treat some categories of information as being more sensitive than others or even the extent to which some kinds of data might count as personal data in the first place.

A fair amount of the discussion here is framed around what kinds of information might be seen by whom (see, for instance, Bellotti & Sellen, 1993; Lederer et al, 2003). Other studies look at what kinds of difference it makes when information is associated with particular people or, by contrast, anonymised or aggregated across numerous individuals (Ackerman et al, 1999). Olson et al (2005) also note how the setting in which different kinds of information is collected can have an impact upon the degree to which it is judged to be sensitive.

Information sharing in the workplace, for instance, is often actively encouraged to improve collaboration and distributed engagements.

For specific kinds of information Olson et al (op cit), based on survey ratings, indicate the two extremes of information sensitivity to be 'transgressions' (e.g. viewing erotic material, personal browsing when it is against company policy) at the level of the most sensitive and work email and work telephone numbers at the least. They also report levels of agreement amongst their participants and find almost complete agreement about:

- Always sharing one's work email and work phone number with one's spouse and coworkers
- Always sharing one's home phone number with one's spouse and children (but not always with co-workers)
- Never giving the credit card number to the public.

(Olson et al, 2005)

By contrast, the greatest levels of *disagreement* related to the following:

- personal items being shared with co-workers
- sharing one's age with a competitor
- [sharing] one's pregnancy status with other team members
- [sharing] one's marital status in a company newsletter
- sharing one's credit card number with one's parents or grandparents
- [sharing] one's pregnancy status with a sibling
- [sharing] work-related documents with family members

(Olson et al, 2005)

Ackerman et al (1999) (see also Cranor et al, 1999) engaged in a similar exercise and found people were most willing to:

- share information about their own preferences, including their favorite television show and favorite snack food
- provide their email address
- provide their age
- provide information about their computer

(Ackerman et al, 1999)

They found about a 50/50 split regarding people giving their full name or postal address. Information that was deemed sensitive by the majority of people included:

- health
- income
- home phone number
- credit card number

- social security number

(Ackerman et al, 1999)

As can be seen from the above, a lot of the time there is a connection made between who can see information of various kinds and the associated sensitivity of that information. Olson et al (2005), for instance, found variation regarding who was prepared to share what with their spouse. They also suggest that, with certain exceptions, there is a lot of overlap regarding what kinds of information might be shared with family members and what might be shared with managers and 'trusted co-workers'. Olson et al base their analysis upon four principal groups of people with whom information might or might not be shared - the 'public'; co-workers; 'family'; and 'spouses' – though they do not drill into these categories in much more depth (beyond 'manager', 'close colleague', 'remote colleague') and do not tackle matters of situational variation beyond basic 'at work' / 'not at work' scenarios. Some researchers are even more general in their treatment of sensitivity. Cadiz and Gupta (2001), for instance, observe that people are generally open to sharing information, except with 'strangers'.

One other strand of research in this vein takes as its focus relative sensitivity of information with regard to 'social norms' and what other people might or might not deem 'normal' behaviour. Hawkey and Inkpen (2006), Huberman et al (2005) and Patil et al (2012) suggest that the most sensitive information is that which people consider to be most divergent from what people might otherwise expect of them. Mao et al (2011) quite specifically identify topics that might fall within this kind of anticipated sensitivity:

- "Sexuality – revelation of sexual orientation or sexual activities and desires
- Expressed Emotions – expression of love/hate for somebody, or emotional outbursts about self
- Confessions – revelation of personal affairs about self or others
- Disrespectful Behaviors – rants and embarrassing behaviors
- Bodily Harm – some accident or adverse reaction (e.g. I just fell down the stairs, hitting my head really bad)
- Illegal Activities – drunk driving or other illegal activities"

(Mao et al, 2011)

#### *2.1.1.6 SITUATED REASONING ABOUT PRIVACY*

The above arguments tend to assume that people's orientations to different categories of information and different kinds of people are relatively static. However, some research has probed this view more deeply and pointed out that just how people may understand some particular piece of information as being sensitive or not is, in fact, a hugely situated affair that cannot be divorced from the circumstances in which that reasoning is encountered.

Hawkey and Inkpen (2006), looking at people's privacy concerns in the context of witnessed online browsing and what they term 'incidental eavesdropping' ("information that can be glanced from casually viewing the screen of a user or overhearing a conversation"), are heading in this direction when they propose 'four dimensions that directly impact the privacy comfort level in a given situation'. The dimensions they have in mind are: 'the user's inherent privacy concerns', 'their level of control' (especially over input devices such as the mouse and the keyboard), 'their relationship to the viewer of the display', and 'the sensitivity of

potentially visible content'. They note that everyone has their browsing activity overseen from time to time but that the most common viewers are people who are relatively trusted such as family and friends. They do also note that colleagues see browser activity quite often and that these may be less comfortable situations. However, the focus in their work is still upon trying to arrive at an encompassing framework rather than a more specific analysis of situated reasoning and the actual sensitivity of the information is largely taken as being fixed and given. Some potentially useful associated features that they do identify include 'recent browsing activity', 'browser settings', 'bookmarks', actions already undertaken to limit what might be witnessable, the 'location of the activity', and the 'type of computer':

> Browsing activities may depend on the device being used and also the location of the browsing. For example, someone with both a home and a work computer may refrain from conducting many personal activities while at work, while someone with access only at work may conduct a broader range of activities. Those using a shared computer without a separate login may not conduct the same activities as those with their own PC or own login. A laptop user may perform the majority of their browsing activities on their laptop and move between locations.
>
> (Hawkey & Inkpen. 2006)

Additionally, Hawkey and Inkpen do acknowledge that "within a viewer category there may be several levels of trust and sharing which may fluctuate depending on recent interpersonal interactions" and that "the impact of potential viewers was highly individual". Dourish et al (2004) undertook a similar kind of study of incidental privacy in the workplace and found, additionally, that people were employing "subtle practices to achieve privacy and security goals, such as positioning a computer screen such that visitors in an office could not see it, or stacking papers according to a secret rationale." Kaasten et al (2002) and Hawkey and Inkpen (2006) note how some recent browser 'enhancements', such as thumbnails for pages recently visited, etc., may actually serve to make privacy management even more difficult, despite offering convenience in other respects. Klasnja et al (2009) focused instead upon how people tried to preserve their privacy when using devices in public places and observed that they used strategies such as "tilting or dimming the screen, or finding a seat against the wall". In the context of handling privacy whilst on the move Holone and Herstad (2010) look at how disabled users use location-based applications and find that there is a sophisticated situated reasoning involved in understanding who might make use of a specific feature of information about their location at a specific time and for the purposes of what when they make decisions about whether to share location information or not.

Whilst it may be clear that there are certain kinds of information that may be exposed to other people at certain times and in certain ways that could be considered problematic, something that research in this area has also made visible is that there is not any one kind of information that one can simply say is always personal and always to be considered private regardless of circumstance. Detailed observations of real-world practice reveal that people actually demonstrate fine-tuned understandings of what might or might not count as personal or private according to different kinds of situations. So, the research we have reported so far would indicate that some information is generally oriented to as being highly sensitive (for instance looking at erotic websites), whilst other things are often treated as wholly inconsequential (for instance work email addresses) and managing such distinctions often involves the pre-specification of security levels (Al-Fedaghi, 2007). However, this overlooks

the highly contingent ways people reason about such things. Previous research regarding how personal information is handled by human agents revealed the real extent to which practices of disclosure and associated reasoning are situated in the specific setting in which they unfold (Tolmie, 2010). In this work a range of disclosure practices were uncovered, including: the management and restriction of information about people through glossing information in highly specific ways; apparent revelations being used to accomplish something quite different such as teasing, criticism or even insult, depending upon the exact character of the cohort present; the management of how information was revealed being tightly bound up with exactly who was being interacted with; revelations being conducted under the auspices of what have previously been called 'fragile stories' (Sacks, 1992), where delicate matters about which people might be criticised are not told to people unless the teller knows the other person might be open to similar kinds of criticism; and the very ways in which information is revealed serving to make manifest to others as well the degree to which it is sensitive (whispering, openly restricting who can see it, passing things over in guarded ways, and so on). All of this was found to indicate that any external definition of what might count as personal or sensitive in any absolute sense is probably beside the point. What counts as personal or sensitive is what people make visibly personal or sensitive by how they orient to its revelation to other people. If they gloss it, if they treat it as only being suitable for certain people, if they deliberately withhold its revelation for some time, if they deliver it covertly or to places where only certain other people will have access, it is, for the people involved, personal or sensitive and can therefore be treated accordingly.

Confronted with this, there is a need to further reflect upon the generic discussions of privacy that abound in both 'expert' discourse and everyday talk. As the above observations make clear, just what makes something private or otherwise is a massively situated concern: just what, just who, just where, just when, and, above all, just *how* makes all the difference. So questions about whether privacy is desirable for x, y or z reason may often provoke a general form of assent, but in situ all kinds of exceptions can be found both as a rationale for exceptional access and as a post hoc rationalisation of why particular access was granted. This is not a matter of people being duplicitous or hypocritical. There is a general sense of 'I don't want people sticking their nose in my business', so to speak. It can even be specified in terms of 'the government', 'the bank', 'social workers', and so on. But these are still categories of person rather than the specific individuals with whom you are interacting here and now. The interactional circumstances and understanding of what you are doing here and now are everything, with no time out. So general notions of privacy are at best a resource that might inform some of these interactions, but people still somehow have to pull off accountably appropriate interaction here and now, with specific recognisable people stood in front of them (literally or through digital mediation). General expectations therefore might be said to amount to a hope that situated reasoning and one's generic view are never put in conflict in the first place, but this makes no allowance for the kinds of contingent and situated action that might actually occur

A further issue here is that, whilst people already have well-honed methods for managing the disclosure of information in their everyday lives and in face-to-face interaction, the resources available to people for doing the same online are thin by comparison. How people reason about their personal information is not something that is just re-written by technology. Instead people try to accommodate technology within the reasoning and expectations they already bring to such matters, with varying degrees of success. Some studies have already tried to

uncover the character of people's practices for disclosure (Kwan & Skoric, 2013; Sleeper et al, 2013; Wang et al, 2011), but very few directly explore and compare these practices in relation to practices of disclosure online. Where comparisons are made (see, for instance, Holone & Herstad, 2010) there is a tendency for the difficulties of shifting from existing practice to digitally-mediated practice to be framed around the evident issues without the active ways in which people already endeavour to bring digital practice in line with other more established practices being made explicit. This is visible in the following observations about the import of technological change for privacy regulation:

> "Technology disrupts privacy regulation in a myriad of ways. Principally, technology lifts or changes environmental affordances and constraints for interactivity so that privacy regulating behaviors fail or are compromised. The changes affect the signalling and perception of situational privacy cues, causing interactions to become decontextualised in time, space, and privacy norms. Technology also alters social perception of an individual's action. As a result, technology permits both deliberate and inadvertent privacy violations and prompts apprehension about the presentation of the social self."
>
> (Boyle & Greenberg, 2005)

Despite the clear importance of these remarks, it is also going to be important to understand the assumptions people already make about how they should interact with technology to actively manage matters of privacy if we are going to understand what kinds of resources should be offered in the future to provide more effective support.

## *2.2 UNDERSTANDING THE SOCIAL CONSTITUTION OF PERSONAL DATA*

The situated nature of how people may reason about their data and not only its sensitivity but also what that data 'means' may actually present some interesting challenges to would-be users of that data, challenges that should not be under-estimated. Whilst concerns about the assembly of digital information in ways that might threaten people's privacy are understandable, given the range of ways in which information can now be captured, a counter-issue to be borne in mind is the amount of work involved in actually assembling information in that way and then somehow making it intelligible to would-be users. Setting aside for a moment the above observations regarding situated reasoning about sensitivity, the potential usability of things like addresses and bank account numbers and images is relatively evident prior to any stipulation regarding what that use might be. However, the as-is usability of some of the 'new' kinds of data about users and their context being postulated in more recent debates, for instance information being gathered by sensors or through the assembly of traces of digital interactions of various kinds, is far less evident. Matters of ambiguity, comprehensibility and the diversity of content are all relevant here. Additionally, other research has emphasized the extent to which much of the data being referred to here requires explication by the people from whom it is gathered before it can make any proper sense. Log data, for instance, is often almost impossible to understand without understanding the exact *human* context within which it was gathered.

To elaborate upon the preceding point, a recent (and as yet unpublished) ethnographic study examined how people managed and shared a range of personal data collected in their homes via a combination of Current Cost energy monitors; motion, humidity, light level and temperature sensors; and a Withings Smart Body Analyser which collated data about household members' weight, percentage body fat, heart rate, and CO2 levels. The output from

these sensors was made available to users through both a simple graph-based visualization that showed peaks and troughs at various times of the day, and the similar graph-based visualization that is provided to all users of Withings devices. A key feature of the study was an emphasis upon examining how people actually reasoned about the data being produced by the sensors as they encountered it. Three key issues were uncovered: the visibility to sensors of the *social* constitution of the data; the recognisability for sensors of the social organisation of the home within which personal data is encountered; and the socially constituted reasoning involved in making this kind of data intelligible.

### *2.2.1 VISIBILITY*

#### *2.2.1.1 THE VISIBILITY OF ACTION*

One of the things noted is that, for actual inhabitants of a home, activities within the home environment are available in a variety of ways, many of which involve reasoning beyond what is literally visible. An example here was the presence of toothpaste and a toothbrush in a particular place in a boy's bedroom. The specific placement of these items was able to make available to his parents on a daily basis whether he had or hadn't brushed his teeth that day. Now what is literally visible here is the actual toothbrush and tube of toothpaste in a specific location. And this, of course, is routinely visible to anyone who might enter his room, in other words, other members of the household. Furthermore, as objects of actual interest the placement of these objects and their actual use are accountably of interest quite specifically for the boy himself and his parents. That is, no-one would think to question why they might care about these things but if anyone else, for instance his sister or a visiting friend, were to inquire about them, it would be perfectly reasonable for the boy or his parents to say 'Why? What's your interest in them?'

For those who do have an interest, the placement of these objects is rendered accountable in terms of their local understanding of the environment and its organisation as a social matter (rather than a technical one). So the kinds of reasoning the boy and his parents might reasonably engage in here are concerns such as: why these things should be there in the first place; what their exact placement might have to say about whether brushing has happened any time recently; and how whether the brushing has taken place is constituted as a moral, accountable, and implicative feature of the organisation of the home. That is, there is a moral understanding that children should brush their teeth and it is the job of parents (not just anyone) to remind them and call them to account for not doing so. To understand the particular moral order in play here one only needs to consider: how would it be if it was the father's toothbrush we were talking about rather than the boy's? Who would then have the right to do the reminding? So the placement of these objects has implications quite specifically for whether the boy has to answer to his parents for apparently not brushing his teeth.

So there is a lot of, often completely taken-for-granted, reasoning wrapped into the positioning and movement of even the most mundane of objects in people's homes. And, the thing is, the actual displacement of things is not by any means always witnessed by all of the members of the household all of the time. Yet, even so, they are still perfectly capable of engaging in this reasoning. So what we can pull out of this example (and countless others like it) is that the visibility of objects in certain places in a known-in-common and and mutually organised space provides for the visibility of action, even though the action might (in literal

terms) be said to have never been seen. The 'visibility' of the action is arrived at by people reasoning about how objects speak of actions as a part of the ordinary social organisation of the home. We can further observe that what the positioning of objects can be seen to 'mean' *really* is consequential for members of the household in equally socially constituted ways.

The question to pose here is: how much of that understanding is available if you can only see the objects and their current (measured) state?

#### *2.2.1.2 ROUTINE AWARENESS*

Another element of the visibility problem is that people are aware of a variety of features of their home as a 'matter of routine'. Here is an example:

> Connie notices that there are regular temperature drops in the kitchen on certain days.
> *Connie: The reason that leaving doors open is another thing that affects the temperature readings. Andrew leaves the garage door open when he's working out there. When I'm not there to keep shutting it".*
> *Andrew 'forgets doors are on hinges, born in a barn really'.*
> *Andrew agrees that it has a really big effect. Andrew plans to change his behaviour but then suggests an adaptation – adding hinges to the garage door because it was a fire door.*
> *"I'll have you house trained in no time on this, this is great."*

In this particular example one of the two inhabitants of the home, Connie, is inspecting the output from the temperature sensors in her kitchen. She knows as a complete matter of routine that her husband, Andrew, likes to have the door to the garage open when he's working in there. The door to the garage opens onto a short corridor that itself leads directly into the kitchen where the door is always open. She also knows as a matter of routine the times and days Andrew particularly likes to work in the garage. Looking at the output from the sensors she can see that these two things coincide and this is enough to prompt her account for what is causing the temperature drop. Indeed, feeling a temperature drop alone, elsewhere in the house, at those times, was frequently enough to provide for this account as a reasonable account.

Despite the fact that there are often quite sophisticated understandings involved in how people are aware of one another's routines and use them as a resource for reasoning about what specific phenomena might amount to, an equally important part of them *being routine* is that they would never dream of making them an object of comment or concern (see Tolmie et al, 2002 for a fuller exposition of this point). So, what is taken to be worthy of remark in the above example is not the business of Andrew working in the garage (though it might be if he suddenly did it in the middle of the night). Instead the observation about the routine here makes available other possible objects of discussion, such as the impact his routine is having on the temperature in the house, the implications of this for the consumption of energy, the need for him to remember to shut intermediate doors when he's going to and fro, and having to shut doors for him. All of this is made recognisable through and premised upon a sensed drop in temperature in the kitchen. But the sensor only 'knows' about changes in temperature over time in a certain location. Everything else hinges upon an awareness of one another's routines and how they feature within the social organisation of the home.

#### *2.2.1.3 WHAT CAN SENSORS SEE?*

What the above materials begin to demonstrate is that sensors can only 'see' very specific things. To underscore the point, here is another example. In this case there is a lot of unusual motion activity in an upstairs bathroom during the day over a two-day period:

> 28/29th December lots of motion activity – when Samantha and family [their children and grandchildren] were all there.
> Connie reasons that all the other bathrooms were probably being used so she might have gone upstairs instead.
> Connie: "I might have used it"
> Andrew: "To run away from them"
> Connie: "Yeah, probably yeah (laughs) …. taking refuge"
> Normally she would have gone to the downstairs bathroom.
> Also, she was cooking a lot.
> Connie: "I took a shower around 11o'clock that day as I had come down in the morning to cook. Then I got changed also so I was using the bathroom a lot that day."

Here you can see one of the inhabitants of this home reasoning about the possible causes of the motion data. It was the period just after Christmas when the family were all there and the downstairs bathrooms were continually in use so it would have made sense to go upstairs instead. Furthermore, she reasons she was needing to use the bathroom more often because having the family all there also meant she was cooking a lot more.

It is important to note here just what it is the sensors are able to 'see' and, more importanly, what that seeing amounts to. What the sensors were able to see was that there was motion. But what the sensors *can't* see is the social order of the home as a reasoned production of its inhabitants. The account of bathroom use as a reasoned and reason-*able* production by Connie here is something that is utterly beyond the scope of the sensors. Thus, a key lesson to take away from these studies is that sensors are fundamentally unable to see the *social* organisation of the home and make sense of what they see as a feature of that organisation. In fact, the central problem here is not just about *seeing* per se, but rather about *seeing as* or *recognition*. For members of households the phenomena sensors capture are about more than just physical phenomena open to being rendered in a purely physical description. They are phenomena to be recognised as an ordinary and accountable part of everyday life. This is something we shall elaborate on further in the following section.

### *2.2.2 RECOGNISABILITY: BEING A MEMBER*

What the above materials have begun to uncover is something in a sense rather unusual. In order to come up with accounts for sensor data looking the way it looks, members of households are being asked to give voice to an order that is usually taken for granted rather than spoken of amongst themselves. In this sense the need to actively engage with this kind of information 'breaches' (see Garfinkel, 1967, for a full exposition of what this might mean) the ordinary order of the home and obliges members to account for it in some way. In fact this taken for granted order is a massively present backdrop to how the home is organised. It is rarely articulated but it is an ever-present resource that informs how members understand 'what is going on really'. The taken for granted character of this order is something that poses significant problems for systems to be able to recognise the import of the data they are gathering. In fact, it is often the case that nothing short of being a member of a household will

be sufficient for recognition of 'what is going on really' to be possible. In this section we shall examine several ways in which this cuts, from the topological organisation of the home as a physical environment to the moral character of how things should be and what might therefore stand in need of explanation. This is terminated with an example that underscores just how much recognition turns upon membership.

#### *2.2.2.1 THE TOPOLOGICAL ORGANISATION OF THE HOME*

An important aspect of how people reason about what is going on in their homes is bound up with how they organise the spatial arrangement of features within them. In particular, the placement of objects within those spatial arrangements at certain times and in certain ways is replete with assumptions about the ordinary order of the space they inhabit, assumptions that both the placement itself, and subsequent understanding of what is being accomplished by that placement, turn upon.

In one of the homes studied there was a counter with some seats at it situated between the kitchen and the sitting room. Often, in the evening a range of books and effects related to the 17 year-old girl who lived in the house with her parents doing her homework could be found situated on this counter. The girl had a desk and workspace in her bedroom but routinely positioned herself at the counter instead. Her own account for doing this was that it was to avoid the distractions of the television and computer in her bedroom, which would lead to her doing something other than her homework. At the same time, there are several aspects of this that provide for member-based reasoning. Even though her books etc. are not strictly 'in place' when they are on the counter, they are accepted as being appropriately positioned there without question at this particular time of the day (though for them to be found there in the morning, on a Sunday afternoon, or whatever, might well prompt others to call her to account for leaving them there). Their placement here at this time enables her parents to do things like see at-a-glance (and even if she's not physically there) that she is in the process of doing her homework (which is itself an appropriate thing for her to be doing, of course). More than this, it also enables her parents to do things like 'knowing that her homework has been done'; even though no explicit question has been raised. Were she to do her homework in her bedroom it might well be something asked of her but with this arrangement the question would be redundant.

Here is another example:

> Working through the house looking at the location of various sensors and how different spaces are used. Leaving the dining room we head towards the garage. As we get near the garage we notice there is a car hoover on the floor outside the garage door.
> *Researcher: So the car hoover lives there.*
> *Andrew: No, no. Not at all. It belongs in the garage but Connie can't reach the shelf, so she's left it there until I get round to dealing with it.*

In this case a car hoover has been 'left out'. But notice that it is not just left out anywhere but rather in a place that is meaningful and implicative for what should happen with it next. By putting it here the hoover itself is instructive to Andrew as to what he should be doing with it and, as members of the same household, there is no need for the instruction to be made explicit.

The important thing with regard to the topological organisation of the home is that it is not so much about precise locations as about placements within the logic of how the home is organised and how people engage in activities within it. Critically, without membership just seeing those arrangements provides no sense of how they might be a reasoned production. These are not absolute logics but rather positionings that are situatedly intelligible to members of the household, so it often takes specifically local membership to see the reasoning within these placements. Others might recognise some of the reasoning around some kinds of placements, for instance the positioning of washing baskets to hold dirty washing, but even this is rarely enough for an understanding of the specific local practices relating to just how these placements are taken to be meaningful for particular members of the household. Certainly, third parties would have trouble, without explanation, seeing the rationality behind the placement of homework books on the kitchen counter or the hoover outside the garage door.

#### *2.2.2.2 THE MORAL AND SOCIAL ORGANISATION OF THE HOME*

Another important aspect of how membership of a specific household informs how people understand what is going on and interact with one another is the routine assumptions they make about one another's rights and responsibilities and potential accountabilities. This sense of the *moral* order of the home[2] provides for an understanding of what features might or might not be taken to be remarkable or problematic. In Deliverable 5.1 we already took note of this moral order and its importance through examples such as the following:

> N: " when he has visitors over .. I am always checking up with him to see what it is he is doing .. when he is on the web ... I get a bit worried... its mainly those games .. with him and his mates.. but as he gets older I more worried about him looking up dodgy stuff ... I can deal with it if he does it,,, but I don't want ...his mates coming round here to look up porn
>
> (UCN, 2014: D5.1)

However, drawing again on the ethnographic studies of sensor use in home environments, we now wish to drill into this matter a little further. Note, for instance, the following example:

> Frank and Susannah have two school age children and spend most of their time at work as an academic and a schoolteacher respectively. They are looking together through the light and humidty sensor data for their bathroom when they happen upon a sudden peak in activity in the middle of the day on a weekday:
> *Susannah: Oooh, what did you do?*
> *Frank: I didn't do anything*
> *Susannah: You did. At 12 o'clock. Look at that*
> *Frank: Where? Nothing*
> *Susannah: No, here*
> *Frank: I could have been up late cos I've had this headache thing… So that's probably me getting up late isn't it? Having a late shower . It's high for a long time… I don't have that long a shower.*
> *Susannah: Yeah, but you could have had a shower and then you could have had a shave.*

[2] Garfinkel (1967) proposes, in fact, that the social order in its totality is a moral order.

These kinds of instances expose quite neatly the assumptions people make regarding what different members of the household should or shouldn't be doing at different times of the day. So, in this case, Frank using the bathroom for a lengthy period of time in the middle of the day is taken not to be a routine feature of how this particular household organises its affairs. The fact that he might have been doing so is therefore made an object of specific comment. Note how Frank is then obliged to provide a reasonable and appropriate account for this course of action. However, the account has to work within the understood moral and social organisation of this specific household. In another household we studied showering in the middle of the day wouldn't even have been worthy of remark. Furthermore, note that all the sensors revealed here was a certain period of increased light and humidity in the bathroom at a specific time of day. Without membership of this specific household it would be hard for anyone to engage in the kind of moral reasoning being exhibited here and the data on its own would offer no kind of insight.

What all of this leads us to is the point we made at the outset of the section: being able to say in any particular household what is in the ordinary way of things, or remarkable, or problematic is something that turns upon membership of the household. It takes membership variously to see: a) what might or might not count as being on the one hand ordinary or on the other hand exceptional; b) just what it would take for something potentially remarkable to be rendered intelligible and reasonable for anyone who is living just here; and c) just what an *appropriate* account would need to look like for this intelligibility to be appealed to. Note how, by the same token, it would also take the same kind of membership for someone to see what might really count as being truly exceptional. By way of an example, in the very same household we have just discussed, unallocated readings were found on several occasions on the Withings scales that had been deployed. These could not be associated to anyone in the household, but it was evident from the weight that the person involved had to be a child. The 'ordinary' account for this they put forward was that it must have been one of the friends of their children using them. However, one of the members of the household who was inclined to believe in such things had an alternative, 'exceptional' explanation: that it was a 'ghost child' and that the house was actually haunted.

As the background reasoning we've been discussing here is something that sensors cannot see it seems inevitable that the reasoning will have to be articulated in other ways that stand above and beyond just the data. In other words, there is a need for local inhabitants to somehow explain what the data really means. The difficulty with this proposition is that this order of reasoning is rarely articulated because it is taken to be just plain for anyone to see. The moral assumptions around which people organise the sociality of their homes form a rarely explicated backdrop to the things that are actually called to account. Uncovering these as an explicit articulation of what it takes to be a member 'around here' is exceedingly difficult and would stand outside of the routine activity of the home.

In order to underscore the point we are making here we are going to draw upon an example first published in 2002 and republished as part of book chapter in 2011:

**The Knock on the Door**

As a part of the studies being conducted under the auspices of the MIME project I had been following the everyday routine of one particular family for some time. There were certain features of this family's routines that had especially piqued our interest.

One of the things that had got us especially excited about this family was a certain arrangement the mother had set up with her next-door neighbour involving knocking on each other's doors. Both the mother in this family and her next-door neighbour had children at the same school and would be setting out to pick up their children from school at the same time each day. The school was close enough that they nearly always walked.

The first time I saw the knock on the door happen it was about three o'clock in the afternoon. The mother I was watching had been sitting out in her garden reading in the sunshine. She looked at her watch then went into her house and locked her back door and started going around shutting windows. Then she headed up her hallway towards the living room and at that moment there was a knock on the door.

When this happened she opened the door a fraction but then, instead of opening the door properly, she went into her living room, continuing to gather bits and pieces together. When she finally did go out of the door the person who had knocked on it, her next-door neighbour, was already walking off up the road.

Now, of course, this happening piqued my curiosity, because it's rather unusual for someone to knock on a door and walk away without waiting for an answer. Typically knocks can achieve a number of things such as being a summons to the people inside, or a way of checking if some room is empty. Walking away without waiting for someone to answer a knock on a door is therefore something that is typically considered to be rather rude. In fact, there is a game of dare played by British children called 'knock down ginger' where this is just how the game proceeds with the goal being to get away unseen and uncaught by the irritated householder who has been needlessly brought to their door.

At the same time, it's pretty unusual to answer a knock by only partly opening the door and then walking away. This certainly isn't the way people normally deal with a summons because it offers no scope for engagement with the person who has summoned you. As the study progressed it became clear this was no chance happening I had witnessed but something systematic because it happened day after day in the same fashion.

It transpired that neither of them had ever discussed this arrangement but had just kind of fallen into it as a way of telling one another they were setting off to school now, so that they might walk together rather than separately. In these circumstances then, the knock on the door was just enough to say 'I'm about to leave', and the half opening of the door was just enough to acknowledge that announcement of imminent departure. They had honed it all down to the barest minimum so that, with beautiful economy, they could bring about a particular co-ordinated routine between their respective households. And within this they understood each other's accountability such that they didn't need to explain why they'd walked off after knocking on the door, or only half-opened it.

Now, of course, it needs to be recognised that this arrangement only applies at around that time of day on a school day. It doesn't take much to see that, were they to do the same at some other time of day or at the weekend they would certainly be quizzed as to why they should do such a thing. Indeed, it was perfectly visible over the course of the study that, for the mother I was watching, a knock on her front door was taken quite expressly to amount to a summons and she always opened her door fully to see who might be there. So one can see how this knock on the door activity is only a thing that is intelligible at a specific time on specific days and that they are mutually oriented to this local and precise intelligibility. Furthermore … it should be noted that this orientation was most especially seeable for someone else like me as I watched them because not once did they pause to remark upon the oddness of it all or problematise it in any kind of way. Yet it was wholly remarkable to me

because it breached all of the ordinary ways in which I might have expected knocks on doors to be handled. In fact, the only time the parties involved saw fit to comment upon the practice in any way was the time I set about trying to capture it on video.

The problem, of course, was that I had always been inside the house with just one of the two parties when this thing had happened and I was rather keen that I be able to capture film of it so that I could show my colleagues the phenomenon 'in the flesh' so to speak. It is a testimony to the foolhardiness of ethnographers that I thought I could simply drive one and a half hours down the road and park outside their houses on a school day before three o'clock and thereby capture it. Nonetheless, that was what I set out to do. It is not until you are sat over the road from someone's house, filming it through a car window, that you start to reflect upon just who might be engaged in the more remarkable behaviour. Clearly, for everyone else passing by I was an object of suspicion. Worse still, as I sat there waiting I started to be plagued by nagging doubts. "Surely", I reasoned, when one considered the odds, there just had to be occasions when rather than knocking on one another's doors they both come out of their houses more or less simultaneously. Furthermore, this was now the last day before the schools' summer holidays so quite probably my only chance.

Leaving to walk to school

"That was good timing"

Of course, as you can see above, this was indeed one of those occasions where they both exited their houses at the same time. As the mother I had been observing came down the path she commented to her neighbour "That was good timing". This was the only time they ever said a word about what they were doing. And this time something quite different had occurred. However, as I recovered from my inevitable disappointment I began to see that what I'd captured was, if anything, even more fortuitous. I realised that the remark 'that was good timing' is not a remark upon the routine itself but rather upon the perfection of its realisation. The beauty of it was that, in these circumstances, the very need for the knock on the door had simply faded away and one could see that it was never simply about knocking on one another's doors at all. Rather the knock on the door was a resource to bring off what they were really after all along, which is to walk to school together rather than alone.

(Tolmie, 2011)

The point of this rather lengthy vignette is to make visible how completely routine arrangements in homes, like the knock on the door that was just described, can sometimes be utterly mysterious to those who are not members of the household. The example also indicates how the householder in question had never really needed to account for this practice before. Rather it was something they had fallen into doing and now took for granted. Thus it took the *social* competence of the ethnographer to recognise this as something that might in any sense count as being remarkable and in need of explanation. But computing systems are not possessed of this kind of social competence. So how are they to recognise what is or isn't in need of explanation in the first place?

### *2.2.3 INTELLIGIBILITY AND THE WORK OF ARTICULATION*

In this final section of our inspection of the *social* constitution of data as personal data, we are going to take a look at a further feature of the studies conducted around this topic. Here our interest is in how, appreciating that the raw data and graphs displaying activity provided little sense of what was actually being revealed about their lives, the inhabitants themselves undertook work to account for the data and make it intelligible to other people around them. To do this we are going to look at both their sense of the gap and the exact character of the kinds of explanations they engaged in. In particular we shall be exposing the significance of this for the handling of personal or private information by abstract third parties and what it might mean for how those third parties go about making sense of personal information arising from de-contextualised data.

#### *2.2.3.1 THE GAP*

It is evident by now from all of the above material that there is a significant gap between what sensors make available regarding people's activities and how those activities are actually meaningful for the inhabitants of the home. Returning again to one of the examples above:

> *Frank: I could have been up late cos I've had this headache thing… So that's probably me getting up late isn't it? Having a late shower . It's high for a long time… I don't have that long a shower.*
> *Susannah: Yeah, but you could have had a shower and then you could have had a shave.*

So, what this example turned upon was sensors revealing a 'light and humidity event' on a particular day at 11 o'clock in the morning. But note the range of commonsense reasoning applied to making sense of this event, e.g.:

- The matter of getting up late (note this itself was accounted for in terms of having had a headache): *Lateness is bound up with an understanding of what the ordinary routine should be. Getting up at this time might be early in some households.*
- The matter of having a late shower: *This is similarly bound up with local reasoning about what might count as late or early doings of particular activities.*
- The matter of it going on for a 'long' time: *This is bound up with a sense of what an appropriate amount of time to have a shower might be for this person in this household.*
- The fact that he could have followed on from having the shower with having a shave: *This is bound up with an understanding of: a) what things make the bathroom humid; b) what kinds of bathroom activities this person might engage in (e.g. shaving); and c) how that might provide for a reasonable explanation of a phenomenon that demands some kind of an account.*

The reasoning going on here is about making the phenomenon an ordinary feature of their lives. This is really a very mundane (if rather personal) thing that provokes a whole concatenation of shared reasoning to arrive at a coherent account. What the reasoning demonstrates very nicely is the gap between a 'technical event' (as that rendered through the data from the sensors) and a 'social event' (as that being rendered through the various local human accounts). The extent of the reasoning here underscores this gap and just what lies between what is visible in the data to 'just anyone competent to read it' (for instance a

researcher or a third party supplier) and the local reasoning applied to find within the data an explicable and locally intelligible event.

#### *2.2.3.2 THE WORK OF EXPLANATION*

Something we particularly want to take note of in the materials we have been exploring here is the fact that a part of what one gets in the accounts provided by the inhabitants of homes is the work of making the data intelligible, for each other, and for others with an accountable interest (in this case a researcher). Sometimes, as in the following, inhabitants will work quite hard to locate explanations in this way:

> Frank and Susannah normally leave the house at about 7:40 but there is still motion being detected
> Frank: OK, so I left the house and then there's some activity at 12:00.
> Susannah: So this has to be the cats because we're not in.
> Frank: Could be cat activity"
> Susannah: We've got two cats and they would…
> Frank: (interrupting) and we're back at 6 so that's kind if weird
> Susannah: That has to be the cats, it can't be anything else can it?
> Frank: No, I don't think so
> Susannah: Although that first one's longer isn't it? How long in duration is that? About 8 minutes. … so generally you see, my perception of what the cats do when we're not here. Fred, so when I leave, like this morning I took the car down to the garage at 9.00 and I clocked, cos David [their son]'s room's opposite ours, and I thought right, Fred's gone back to bed. So he's asleep on David's bed so and there he will remain, apart from getting up to have a look round for food. He'll get up if there's somebody that comes into the house but otherwise he'll stay asleep. But as I was coming down the stairs, Moomin shot up the stairs, so I thought I bet she's going back to bed. So they'll get up when we get up but then as soon as we leave I do think they go back to bed then have periods when they might get up sometimes…

So what we want to draw attention to here is the fact that effort is always devoted to making what is visible in the data coherent in terms of the ordinary organisation of the home. The 'problem' here is that motion has been detected when no-one is in the house. They do, however, have two cats so the 'reasonable' option is to attribute this motion to the cats (not ghosts). Notice how Susannah in particular takes some time here to render visible in her account just how the cat explanation *is* a reasonable explanation of the data as a meaningful feature of their home. The research reported here found that data that is not readily accounted for in this way is treated as being either suspect (e.g. the sensor must be faulty) or else relating to some hidden process (e.g. systems doing things that you'd have to be a plumber or an electrician to know about). Critically, data is always understood to be somehow speaking to ordinary phenomena: 'ordinary' as in accountable to the social organisation of the home, not 'ordinary' as in accountable to the parameters of measured phenomenon.

Another, more accurate way of describing the work of explanation is 'articulation work':

- It articulates what it is the data might be showing you 'really'
- It articulates the grounds upon which this might then be meaningful

As we have seen above, a great deal of sensor data is nothing without this work of articulation because the data provided by sensors is not self-explicating. Instead, explicating the data takes articulation between things like abstract numbers or graphs and the understood social organisation of the setting it relates to. Articulation work is a methodical way of working between standard representations of times and events and what is known of how events are locally ordered and produced. It is premised upon an assumption that the data is always somehow accountable (even if only as it being the outcome of a fault or whatever). As a matter of method the work of articulation takes finding within generic representations the means of providing a *locally* coherent account.

Here we find another challenge that lies under the surface of a great deal of the more general angst that is exhibited regarding the scope for personal data to be harvested by the new kinds of technologies people are putting in their homes. What falls out of the above discussion is that the issue with not having any control over personal information and it going out of the door in a raw format is that it is being passed on without the opportunity for *any* kind of account at all. Thus it might, in principle, be open to *any* kind of account or understanding by *any* other unknown person. With a known recipient it is always accounted for somehow and the work of accounting for it is the very guts of privacy management as a practical accomplishment because the account is never generic but rather always situated and tailored for the specific recipient of that account.

In sum, the above observations bring to the fore the fact that reasoning about personal information does not just get re-written by technology. It is therefore important to understand reasoning of this order, whether or not technology is involved, and the kinds of practices that have evolved to support such reasoning (Kwan & Skoric, 2013; Sleeper et al, 2013; Wang et al, 2011). One aspect of this, for instance, is that much of the information that is taken to be personal and potentially sensitive is ordinarily completely visible to those with whom one interacts on a daily basis. Another aspect is the amount of local understanding involved in being able to make the data in any sense intelligible. This complicates the picture regarding just how digitally mediated personal information might be made available or subject to reasoning to third parties, willingly or otherwise. This is particularly the case where such rich data types combine to create a 'contextual digital footprint' (Mortier et al, 2010; Sheridan et al, 2011), with residues of such 'footprints' becoming considered as digital legacies with a longevity that expands beyond a particular user's own lifespan (Barua et al, 2011).

## *2.3 UNDERSTANDING THE MACHINE*

In what is almost the reverse of the preceding observations, if it is potentially difficult for would-be users of data about people to understand what the data means, so too is it potentially troublesome for people themselves to see and understand just what systems themselves may be telling them is being done with their data and just what technical descriptions of privacy mechanisms may 'mean'. In fact a number of researchers have pointed to the distinction between ordinary reasoning about privacy and technical understandings of what privacy mechanisms should look like. The gap between these two ways of looking at privacy, and the potential this creates for users to lose sight of what systems may really be doing with their data, has resulted in some researchers putting a particular emphasis upon the need to make available privacy mechanisms intelligible to users if they are going to be effective, e.g.:

> "To participate in meaningful privacy practice in the context of technical systems, people require opportunities to *understand* the extent of the systems' alignment with relevant practice and to conduct discernible social *action* through intuitive or sensible engagement with the system" [original italics]
>
> (Lederer et al, 2004)

In relation to this Lederer et al identify particular ways in which intelligibility can be undermined, e.g. by: 'obscuring potential information flow'; 'obscuring actual information flow'; 'emphasizing configuration over action'; there being a lack of 'coarse-grained control'; and the inhibition of 'existing practice'.

One of the areas where lack of visibility and understanding has received particular interest is in the context of social networking activities. However, whilst acknowledging the need for improved 'defaults' and 'better tools for managing privacy' some researchers (e.g. Liu et al, 2011) have suggested that there is not yet enough information about the exact character of issues surrounding privacy in the context of using social networking sites like Facebook. Liu et al (op cit) found that overall just over a third of all content on Facebook is shared using default privacy settings but that, at the same time, only a third of the privacy settings used "matched users' expectations". In relation to this they found that the usual consequence of this was that information was exposed to 'more users than expected'. They suggest a possible solution here might be to actively use 'user-created friend lists' for the management of privacy, though they do not go into detail about how this might be best accomplished. Studies of Twitter use (Mao et al, 2011) have also indicated that, even where there is already an assumption that everything is public, there is still space for people to not fully grasp the privacy implications of their actions. They show that in a variety of what they term 'leaks' people may inadvertently reveal travel and vacation plans, leaving themselves open to robbery, and may also reveal more than they might wish about medical conditions, as well as composing tweets they subsequently regret whilst under the influence of alcohol.

Other researchers point to poor user understanding of even relatively routine aspects of web browsing such as cookies (Ackerman et al, 1999; Cranor et al, 1999). The range of confusion here is made manifest in the following comments:

> "Cookies can determine my identity from visiting the site"
> "I may have a false sense of security but I understand that as long as I accept 'no cookies' the site managers cannot access my email address and other personal information."
> "A cookie can only provide information I have already given, so what is the harm?"
> "I am not quite sure what a cookie is, but I have an idea."
>
> (Cranor et al, 1999)

Some studies have pointed to the extent to which people are unaware of how a range of wholly ordinary activities on their part may be contributing data to unseen third parties. Blasbalg et al (2012) comment on how something as 'simple as buying goods from a grocery store with a credit card' or buying things online that are promoted through social networking sites can, in fact, facilitate the building of a 'fairly complete picture of a person's lifestyle' (see Petersen, 1995). They also note the range of third parties who might have an interest in purchasing this kind of information, such as insurance companies.

Another area of confusion pointed to (and of potentially particular moment for UCN) is the nature of public wireless networks and their security. Chin et al (2012), for instance, note that the unwitting consequence of providing network security warnings when people are connected to untrusted networks is that 'some users have become afraid of all wireless networks without understanding the threat', which serves to deter use of mobile devices in public places.

Indeed, security settings and associated permissions are reported to be a particular issue for user comprehension (Patil et al, 2013). Here, Felt et al (2012) discovered in a large-scale study of Android users that "only 17% [of their participants] paid attention to the permissions (including ones which grant an application access to privacy-sensitive data) when installing an application". On top of this only 3% "demonstrated full comprehension of the permissions screen". Another study by Kelley et al (2012) also found that most Android users "found it difficult to understand the terms and wording of the Android permissions".

Klasnja et al (2009) point to a security issue to do with the use of web services that is not only not understood but often invisible to users. We have already noted in 2.1.1.4 that many users accept the need to provide certain kinds of information to web sites in order to receive certain kinds of services and it has become relatively routine to provide a name, your age, your post code and certain particular preferences. When signing up to wireless network services this information is often also shared with advertisers and other third parties and in many cases without encryption. Klasnja et al (op cit) comment that:

> "A majority of the large Web-based email services, for example, encrypt the login process, but not the contents of email messages. Anyone along the path between the user and the service's data center could intercept this information, opening users to privacy and security risks."

Klasnja et al point to various other privacy risks associated with the use of wireless networks as well, including fake access points harvesting data, tracking and surveillance of people through their use of networks, 'eavesdropping' on people's transmissions, and so on, without any real way in which users can assess the relative security of the network or who they might be visible to. Indeed, they suggest that people's 'understanding of the risks associated with Wi-Fi use is limited', and that they were often 'not aware that information sent over Wi-Fi could be seen by others', regardless of their use of firewalls and antivirus software (which is frequently understood to be protection against 'all evils').

An associated aspect of the debate here is the extent to which people understand the formulations and policies of online service providers and the impact these may or may not have upon their legal rights. Lahlou et al (2005), for instance, note that people have still largely not caught up with the way in which things like online shopping have changed the rules about what can or cannot be done with their data:

> "In a computer world without sensory borders, rules such as "if I can see you, you can see me" are no longer relevant".
>
> (Lahlou et al, 2005)

Researchers (e.g. Baumer et al, 2003; Berendt et al, 2005) investigating the impact of having highly visible privacy certification and policies, have found that these often make little or no

difference and that trust is frequently founded upon how well-known the name of the web-site provider is instead. Others such as Ackerman et al (1999) have suggested that the impact of visible policies depends upon the 'type' of user, with 'pragmatists' being more swayed by this than others. Berendt et al (2005) note that an additional difficulty with people putting faith in evidence of jurisdiction is that they may exercise less care about the potential risk of disclosure arising from their own activites. This, they suggest, adds impetus to the need to come up with proper software-based solutions to the need for privacy protection. This is echoed by others such as Conti and Sobiesk (2007) who point to a high level of complacency amongst many users who work largely on the assumption that "an honest man has nothing to fear". Klasnja et al (2009) underscored this through their own investigations where attitudes were summed up by the following remark: "I *kind of trust my bank and my credit cards...when they say that this is hacker-proof, that it truly is"*.

However, despite the need for such design investment, others have noted that there is a large inclination on the part of many designers to abrogate their responsibility and assume that privacy is 'someone else's problem', not their's (Blasberg et al, 2012; Lahlou et al, 2005). This view risks putting systems design in a similar position to users regarding a rather sanguine hope that policy and legislation will somehow be enough.

As we shall be stressing again in the next section on requirements, all of this has led to some researchers calling for systems designers to step into the breach and make sure that system handling of data and accounts of its actions be rendered more transparent and intelligible to users so that people can make more informed choices about how to proceed.

# 3 PRIVACY REQUIREMENTS AND SECURITY MODELS FOR UCN

## *3.1 INTRODUCTION*

As stated in the description of work of UCN, users are to be equipped with an intelligent module called a Personal Information Hub (PIH for short) which is in charge of orchestrating the interactions between the different devices present in a house: These devices range from laptops and smart TVs to cameras and sensors that keep track of the activity (suspicious or not) inside a home.

One basic functionality of this PIH is to collect data (such as energy consumption, camera feeds, favorite TV shows) from the electronic devices and sensors in a home. While the collection is not the aim of setting up the PIH, it can be considered as a starting point that will enable the PIH to provide a multifaceted service afterwards. For instance, the data collected from sensors when analyzed and processed properly can be leveraged by the PIH to detect a suspicious presence or behaviour. When such a behaviour is detected, camera feeds can then be used to either confirm or refute the suspicion raised by the sensors (see the use case examples in UCN Deliverable 5.1). In a similar manner, the PIH can be used to alert the user when the energy consumption in the house exceeds some predefined threshold or when the gate to the house is not locked ... etc. Furthermore, if a feedback loop exists between the user and the PIH, the PIH can be used to provide more intelligent services such as TV show recommendations based on what the user has watched so far. The PIH in such scenarios acts as an intermediary between the user and the home appliances, which by interpreting the collected data offers valuable services such as home energy management and local recommendations.

On the other hand, the data collected by several different PIHs can once again be collected by a third party or service in order to derive some meaningful statistics. For example, in the context of smart cities, one trending use case is air pollution measurement whereby the goal is to collect individual measurement from a very large number of PIHs to further compute the city's air quality index. Another prominent application that can exploit data from multiple different PIHs is recommendation systems. By having a large number of different user profiles, recommenders can first identify clusters of users that share the same movie taste for instance and further propose relevant movies or advertisements. In these two previous scenarios, the PIH is considered as an individual source of information from which third parties retrieve useful data.

In line with what was discussed above and the contributions of UCN Deliverable D5.1, we identify two global use cases that will steer the definition of our security models, namely, recommendation systems and smart homes: while in the context of recommendation systems, users are often required to share some data with third parties (recommenders), in most smart home applications, the user is the only party who has a direct access to the data at the PIH (cf. Appendix). We believe that sharing even the smallest amount of (sensitive) data with a third party (that is not necessarily trusted) inherently puts the privacy of the user at risk[3]. This

[3] In this deliverable, we do not take into account the case whereby some adversaries infer sensitive information without having access to the actual data and therefore violate users'

means that to enable the user to take full advantage of the plethora of services that recommenders and smart homes have to offer without undermining his privacy, one should design suitable privacy preserving primitives and protocols that will underpin such services. Ultimately, the goal would be to design solutions that are provably secure and privacy preserving to a certain degree.

This mandates that, as part of the UCN effort to deliver privacy preserving applications, we should first come up with a proper security model under which our solutions will be secure. This security model should consider the threats that a user is faced with once his data is shared and should accordingly define the trust levels that will govern the interactions between involved parties. Yet defining such a security model is not a straightforward task inasmuch as these requirements differ from one application scenario to another. Furthermore, while ideally we strive to satisfy all users' privacy requirements, this can sometimes be at odds with service efficiency and accuracy. Along these lines, we propose in what follows three security models instead of a unique generic one, and which reflect the different privacy requirements that are associated with the use cases of smart homes and recommendations defined in deliverable D5.1. The rationale behind these security models is that, as discussed above, privacy/security requirements are highly context sensitive, and often users are willing to relax them in the hope of getting a better service or perhaps a financial compensation as discussed in previous sections. Given these three security models and the user inclination to share (or not share) data, we could find a compromise between user satisfaction regarding privacy protection and the quality of service in terms of efficiency and accuracy.

## *3.2 UCN ENVIRONMENT AND PRIVACY ASPECTS*

Before moving to the definition of the proposed three security models, we introduce the environment in which UCN applications will operate. First of all, we assume that PIHs are totally trusted by and under full control of the user. The security and privacy issues pertaining to the actual installation of the PIH at home (e.g. who is allowed to access the data on the PIH or change the configuration of PIH) can be mitigated using classical authentication and access control mechanisms that are compatible with the technology underlying PIH.

Additionally, PIH's storage capacities are assumed to be limited. Therefore, these devices cannot store a large amount of data indefinitely. On the other hand, it is advisable that all collected data remain available. Therefore, we believe that PIHs should be allowed to outsource their storage to some cloud servers. Although the outsourcing of the collected data solves the issue of storage scarcity at the PIH, it raises new security and privacy challenges since i) cloud servers are regarded as potentially malicious entities that are interested in learning the content of the outsourced data and ii) the outsourced data is generally of very sensitive nature (camera feeds, energy consumption...etc.) that may give away information about the daily routine inside a household. Thus, the proposed security models will also take this additional party (whenever it is involved) into account. The main issue that one has to deal with when outsourcing data is the confidentiality of data.

---

privacy. "For example, Calandrino et al. (2011) have shown how collaborative filtering systems inherently leak to users information about the transactions of other users."

To address confidentiality the first thing that comes to mind is encrypting the data to be outsourced using the user key. However, encryption when performed using classical techniques such as block ciphers or existing asymmetric encryption limits the operations that the user or any other authorized party can perform on the outsourced data. One example to demonstrate the limitations of classical encryption is that once data is encrypted using the well-known symmetric encryption algorithm, AES, the user can no longer perform an efficient search on his data. The only possibility would be for the user to download his data in its entirety and to decrypt it. Such a naive solution is deemed to be impractical most of the time in terms of both communication and computational complexity. This simple example illustrates that depending on the application one should design some new primitives allowing the privacy preserving computation over the remote data.

Finally, we would like to note that PIHs may also be allowed to directly transmit some data to third parties such as recommenders. Hence some application scenarios may not involve the cloud service provider.

## *3.3 UCN SECURITY MODELS*

The three security models we propose in this deliverable are defined with respect to access rights that a user is willing to grant to third parties: They range from the conservative model in which no third party is allowed to learn any information about the user (**No access**), to a more relaxed model where an authorized third party is allowed to derive in a *controlled fashion* some personal information about the user (**Full access**) (cf. Figure 1). We note that even in the relaxed security model, the user should be able to control what information is divulged to third parties. We also propose a model in between **(Partial access)** in the case where multiple PIHs are involved and, as individual information is not accessible, authorized third parties can get some insight about the user population as a whole only. We describe the main challenges these security models imply for the design of new privacy preserving primitives.

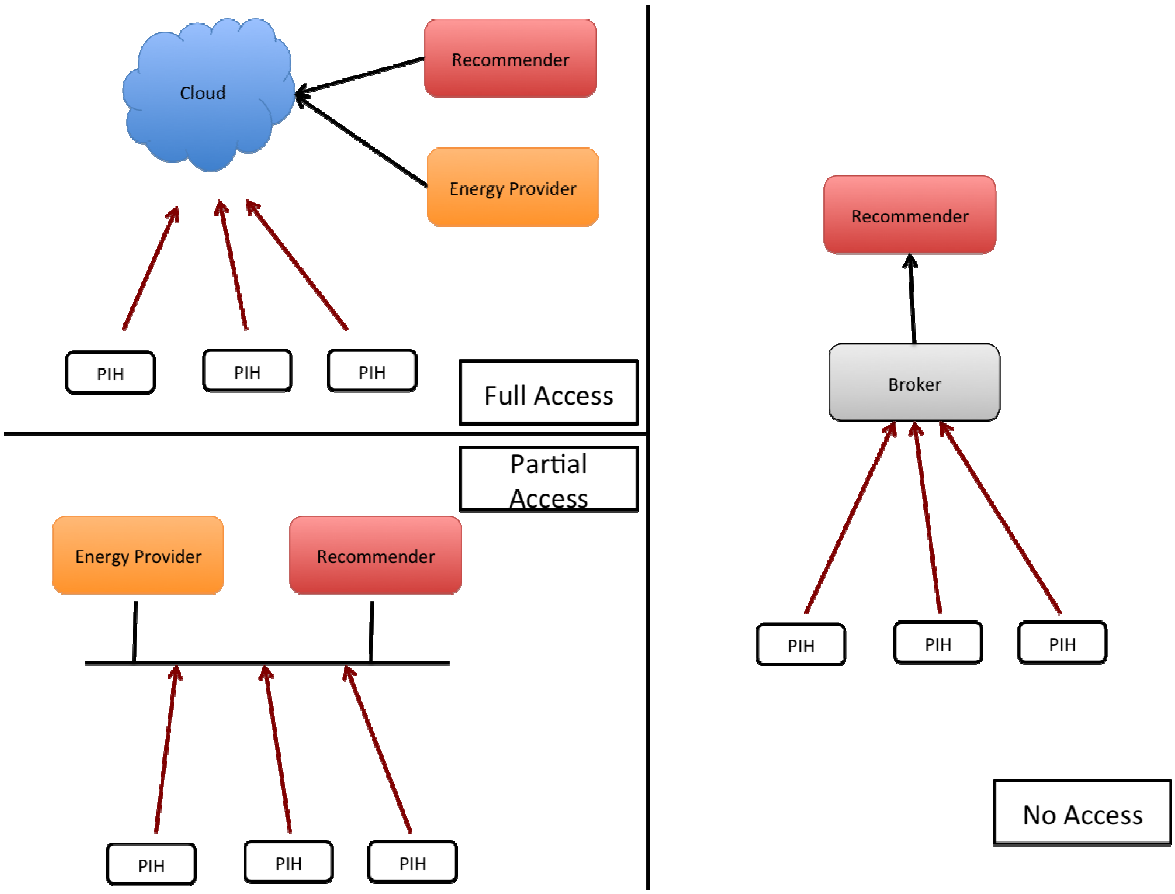


**Figure 1: Access-Related Security Models**

### *3.3.1 NO ACCESS TO USER DATA*

In this first and most strict model users do not want any third party to have access to the cleartext content of data. For example, in the recommendation use case, users may not want to reveal their profiles to recommenders at all; however, they should still be able to anonymously receive relevant recommendations. The goal would be to receive some recommendations without having any direct interaction with recommenders. Such a model can be achieved on the one hand by applying the recommendation algorithms directly at the PIH or on the other hand by using some cryptographic functions named secret matching which enable the secure comparison of profiles with advertisements. In the case of smart home applications, this security model can be applied to scenarios where only the user is allowed to process his data outsourced to the cloud, hence encrypted.

#### *3.3.1.1 RECOMMENDATION AT THE PIH*

Decentralization of personal user data was previously studied in the context of online advertising, which entails gathering browsing and behavioral data. Adnostic (Toubiana et al. 2010) and Privad (Guha et al. 2009) are two privacy-preserving systems, which offer advertising services while storing private user data on the user side. This is achieved by pushing parts of the advertisement selection process to the client. However, this solution is

inapplicable to Matrix Factorization and Collaborative Filtering recommenders, which are too computationally intensive to run on the client. While Vallet et al. (2014) proposed a recommender system where user ratings are stored only on the client side, their solution still requires exposing the data to the recommender, and trusts it not to retain the data.

#### *3.3.1.2 BROKER-BASED RECOMMENDATION*

In order to protect the privacy of such users, one interesting approach is to rely on the existence of an additional party, the Broker, who may be in charge of forwarding recommendations to relevant users. The integration of such a new party should, of course, not harm the privacy of users either: users would not want to reveal their profiles to this broker as well. The broker should only be in charge of comparing users' profiles (interests) received from PIHs with recommenders' advertisements while these two sources of information remain protected. Advertisements should be protected in order for the broker not to discover the interests of users in case there is a match. This third party therefore computes some ratio on the similarity between users' profiles and recommenders' advertisements without having any control or knowledge on the private data it receives.

In Shifka et al (2011) the authors propose a dedicated solution, which ensures the privacy preserving similarity ratio computation thanks to the combination of a searchable encryption scheme (Boneh et al, 2004) with counting bloom filters (Fan et al, 2000). The solution also ensures that each party (the user or the recommender) can verify the correctness of the computed ratio. The authors show that the cost at the Broker is very small and thus the solution is scalable to cover a large number of users. Thanks to this building block, the user will not reveal any information to either the recommenders or the Broker. Additionally, since the broker provides a proof of correctness on the computed ratio, the user does not need to trust this third party even on the computation. The security model is therefore stronger than the classical "honest-but-curious" one.

#### *3.3.1.3 PRIVACY PRESERVING LOOKUP FOR SMART HOMES*

In the smart home use case, sensors and cameras generate some data and the PIH in turn collects and encrypts this data, and finally outsources it to a cloud server at some fixed time intervals. The user on the other hand, is expected to search the outsourced data for some keywords/(keyphrases). For instance, the user who is actually not at home may want to look for the keyphrase "temperature recorded on 25th of June was 22 C". If the search result returns a yes, then the user deduces that no suspicious activity was recorded; otherwise, the user infers that either the AC system was faulty that day or someone has changed its settings. The cloud server in this case is assumed to be semi-honest (honest-but-curious), that is, it is interested in learning the content of the outsourced data and the search queries, but still executes the search protocol correctly. In other words, the cloud server always returns the correct result of the search be it a yes or a no.

Privacy preserving lookup deals with the problem of performing word search over encrypted data. Solutions for privacy preserving lookup vary depending on the targeted application scenarios. For example, solutions where only the user generating the data performs word search cannot be easily customized to fit a setting where data is generated by multiple users who encrypt their generated data using different secret keys, whereas solutions for multi-user settings are too complex for applications where only a single user is involved in the

generation of data and the search operations. Accordingly, we suggest defining the security and privacy model per application scenario. Notably, we describe two application scenarios, one is dedicated to smart homes, whereas the others targets recommender systems, and for each application scenario, we identify a set of requirements and conditions that should be satisfied.

Even though the literature is replete with solutions (Boneh et al, 2004; Waters et al, 2004; Curtmola et al, 2006; Bellare et al, 2007: Kamara et al, 2012; Blass et al, 2012) for privacy preserving lookup solutions that allow a user to search its encrypted data efficiently, most of these proposed solutions (Boneh et al, 2004; Waters et al, 2004; Curtmola et al, 2006; Bellare et al, 2007: Kamara et al, 2012) leak the search result to the cloud server (that is, the word the user is looking for is in the file or not) and the access patterns (the cloud server learns if the user looked for the same word several times or not). While one might argue that the impact of such information leakage is minimal (the cloud server may not be able to make much use of such information without some a priori knowledge), it is still preferable to design conservative solutions such as Blass et al (2012) that in addition to not leaking information about outsourced data, do not disclose information about the content of the search queries and the search results.

### *3.3.2 PARTIAL ACCESS TO USER DATA*

In this security model, the users are inclined to allow third parties (energy providers for example) to process their data to derive some statistics about a given behaviour or trend. Accordingly, third parties need to crawl and mine data from a large number of users so as to get representative and meaningful results. Users however do not want to divulge any personal information unless it is proven to be necessary to the correct computation of the statistics. Ideally, the third parties are only given access to the users' encrypted data but are equipped with some technical means that enable them derive the statistics over encrypted data.

One way to implement such functionality is privacy preserving data aggregation. In principle, privacy preserving data aggregation allows a third party to compute a function over the private inputs of a group of users in such a way that the third party only learns the value of the aggregate (i.e. the function output) and nothing else. That is to say that during the aggregation process, the third party does not and cannot infer any bit of information about the individual inputs of the users involved in the aggregation protocol. Given the above definition, one can safely assume that privacy preserving aggregation encompasses any protocol that aims at computing any function over the users' input as long as this computation is performed in a privacy-preserving manner. Existing solutions that enable the computation of arbitrary (polynomial) functions are not efficient: such solutions in fact require either the use of fully homomorphic encryption or garbled circuits (Gentry, 2009; Gennaro et al, 2010), both of which are computationally demanding. It follows that instead of focusing on generic solutions that support the computation of any functions, the approach in the research community has been to come up with dedicated solutions for specific operations such as: sum, average, linear regression ... etc. (cf. (Shi et al, 2011, Joye & Libert, 2013; Nikolaenko et al, 2013)). These solutions generally build upon the well-established techniques of homomorphic encryption and secret sharing to answer some of the challenges related to data aggregation.

In a similar vein, UCN aims at proposing dedicated and efficient solutions for some aggregate operations (sums, average ... etc.) that can be successfully applied to the following scenarios:

#### *3.3.2.1 DATA AGGREGATION FOR SMART HOMES*

An energy provider wants to assess the energy consumption in a neighbourhood, and hence, will ask the users' permission to collect their energy consumption**.** Users on the other hand, are reluctant to share such information as it may disclose information about their daily routine. A peak in energy consumption could indicate for instance that the household is having guests, whereas low energy consumption could reveal that the house is empty. We note that although the leakage of such information may seem harmless, it is still preferable to design solutions that prevent the energy provider from acquiring such insights. Thus in this scenario, we assume that the energy provider is not trusted, and accordingly is only provided with encrypted inputs from the users. The goal of any aggregation protocol is to allow the energy provider given the encrypted data and some secret key to compute the aggregate value (e.g. sum) in a privacy-preserving manner without revealing the individual value. Another crucial point that should be taken into account when designing such protocols is that the user inputs are collected using sensors which are prone to failure, and consequently, any designed solution for this application scenario should - in addition to being privacy-preserving - be robust against arbitrary sensor failures.

#### *3.3.2.2 DATA AGGREGATION FOR RECOMMENDER SYSTEMS*

Data aggregation could be a useful tool for recommenders that are only interested in learning the behavioral profile of a cluster of users rather than their individual profiles. A recommender in that case is not necessarily in direct contact with users, but still has access to their encrypted data, which might be outsourced to the cloud. By exploiting this encrypted data, the recommender wants to learn for instance the trending topics or content across a group of users. To be able to derive this aggregate profile, the recommender is assumed to offer some sort of financial incentive to users that are part of the cluster (or at least obtain their consent) in exchange for some keying material that will enable it to perform the aggregation correctly. As in the previous scenario, the recommender is not trusted and is only given access to the users' encrypted data. This scenario, however, differs from the previous one in two ways: i) Users in this scenario are not prone to failure since the PIH can safely be assumed to be always connected to Internet; ii) It is easier for users in this scenario to either leave or join the recommender system (due for instance to an unsubscription or new subscription). Hence, a privacy preserving aggregation solution for recommender systems should support users joining and leaving.

### *3.3.3 FULL ACCESS TO USER DATA*

In this model, we consider the case where users are not only willing to share their encrypted data as in the partial access model, but are also willing to allow third parties to infer and learn some useful information about them themselves (for instance: taste in music, favorite TV shows or energy consumption). Of course, this sharing of information should only be performed after the users' explicit consent and the user should always control the amount of data exposed and to whom it is disclosed. The usefulness of such a model can be illustrated by two application scenarios: one dedicated to smart homes, the second aimed at the recommendation system use case. Both scenarios demonstrate that by employing suitable

privacy preserving lookup mechanisms, one can offer a valuable and meaningful service while ensuring that the disclosure of data is totally under the user's control.

#### *3.3.3.1 DELEGATED LOOKUP FOR RECOMMENDER SYSTEMS*

In this application scenario, the querier of data is the recommender who, based on the search results, can learn which contents a given user is interested in. This scenario presumes that it is OK for the recommender to identify the users interested in its content, and accordingly some trust is placed in the recommender. This can be translated to the real world by compelling the recommender to offer some financial compensation to the user (for example: discount on a subscription) in exchange for being granted the right to search the user's outsourced data. The idea behind this application scenario is that the recommender searches some specific user's data to check whether this user watched a particular TV show or a movie, and based on the search results (i.e. the number of hits), the recommender constructs a profile for the targeted user. This profile can be used later to send recommendations or targeted adds to the user.

Note here that while it is important to enable the recommender to search the outsourced encrypted data, it is equally important to ensure that i) the recommender only learns the result of the search and nothing else and that ii) the user is empowered with the capability to revoke recommender access at will (this can be reflected in real life by an unsubscription from the recommender service). These two requirements may be addressed by privacy preserving delegated lookup techniques with efficient revocation (Elkhiyaoui et al, 2014), which in addition to assuring that the cloud server storing the encrypted data learns nothing about the query and the result from the search operations, also guarantees that the recommender learns only the result of its search operations and nothing else, and provides the user with revocation capabilities that do not overburden him computationally.

#### *3.3.3.2 LOOKUP ON MULTI-USER DATA FOR SMART CITIES*

We could also imagine another application scenario in which city officials want to search the data of multiple users to assess the air pollution index throughout the city. To this end, city hall deploys sensors on users' homes and each sensor measures air pollution and transmits data to its corresponding PIH. Later, each users' PIH processes these measurements and stores a bit 0 or 1 indicating whether the air pollution index during the day is under some predefined threshold or not. The different PIHs involved in this procedure are then assumed to outsource this information to the cloud after encryption. The city officials look up, for instance, the key phrase "pollution index 0 on June 25th" in the users' outsourced data. Then, given the result of the search, city officials infer whether the current pollution index is acceptable or not. It is important to indicate here that performing lookup on data generated by multiple users puts forth challenges of key distribution and key management. A naive solution in this multi-user setting would be for city officials to issue a dedicated search query for each user. Such a solution is impractical as the number of keys the recommender is required to keep and the complexity of the search query grow with the number of users in the system. This shows that ideally one should design a solution in which the recommender is able to search the data of multiple users by issuing a single search query using a single key.

## *3.4 ADDITIONAL PRIVACY RISKS*

The UCN security models address the privacy and security risks that result from direct access to user data. However, there are privacy risks that inherently result from the collaborative nature of UCN services, and therefore apply across all the security models. For example, Calandrino et al. (2011) have shown how collaborative filtering systems inherently leak to users information about the transactions of other users. The attack here relies on observing the outputs of the recommender, and does not require any access to user data.

Such leaks can be mitigated by leveraging privacy techniques that protect the output of a computation, for instance differentially private mechanisms (cf. (McSherry and Mironov, 2009), as opposed to techniques that protect the process of computation, such as the use of homomorphic encryption. However, since the use of these techniques typically comes in direct conflict with the accuracy of the computation, their value should be weighed against the cost in utility.

# 4 CONCLUSION

This document has addressed two principal concerns: 1) presenting an in-depth assessment of user orientations to privacy and ethics when using computing systems and the kinds of requirements for design implicated by these orientations; and 2) presenting how the security models being developed in UCN will specifically seek to address user requirements regarding privacy.

For the first part of this we have examined in some detail ordinary everyday understandings of privacy across a variety of different concerns, including: worries about privacy as a general matter but also in terms of quite specific threats; the relationship between those threats and the use of specific devices and applications in specific settings; the inevitable trade-offs people are obliged to make regarding privacy all of the time; just what information might be seen to count as personal in the first place; and the degree to which how privacy is oriented to is a product of the specific situations in which people find themselves rather than being subject to once-and-for-all reasoning. We have also looked at the ways in which personal data, and what might be done with personal data, are features of a broader body of social reasoning. This presents a number of issues regarding both the capture and the intelligibility of data for computing systems. The personal nature of phenomena, it turns out, whilst being evident to people themselves in a particular setting, goes some significant way beyond what might be deemed to be just 'perceptually' visible. It is embedded in local understandings of the organisation of the environment, what might count as routine or remarkable about the features of that environment, and the moral implications of how those features are arranged. None of this is 'just available' to sensors or within system logs. Nor can 'just anyone' see it. This means that the meaning of what is captured requires input from the people it relates to and obtaining this kind of input is a non-trivial problem. This means that what can be deduced from data regarding its private character or otherwise is something that is hard to just build in to computer systems. An additional concern we have pointed to is the fact that much of what systems are doing with people's data is far from apparent to people themselves, making it hard for people to trust that systems are properly oriented to what they might or might not consider to be private. The further outcome of all this is an urgent need for systems to be 'ethical by design'.

Confronted with these various issues the second part of the deliverable has focused upon how the UCN environment will be constituted in order to meet the privacy requirements of users. It has also described how the security models within UCN will be applied across three distinct kinds of situations regarding the availability of user data and in two kinds of global circumstance: the operation of recommendation systems; and the presence of 'smart home' technical systems. The central feature of the UCN solution is the use of Personal Information Hubs (PIHs) that house available user data and that are only direct accessible by the users themselves. The foundational principle here is that, as systems are not in a position to make judgments about the privacy or otherwise of data and users are not inclined to trust machines to make such judgments, the best policy is to not make any data at all automatically available without express user permission. On the back of this principle we have elaborated a variety of ways in which the PIH will operate in order to provide the effective delivery of services whilst maintaining user privacy and security. We have, additionally, indicated ways in which the UCN solution may nonetheless retain certain kinds of vulnerability and how these

vulnerabilities may also be addressed, for instance through the application of differential privacy mechanisms, whilst simultaneously indicating a need to engage in further cost-benefit analyses in order to arrive at the best kind of system configuration.

Subsequent work in this area will examine in greater detail the actual development and application of the models and principles outlined in this document and their outcomes in user environments where they have been deployed. In particular, as details of proposed designs meeting the user requirements emerge, the associated ethical issues are being examined and fed back, iteratively, into the design process. As these concerns are deeply embedded in the details of how personal data is handled, and the interaction between these processes and the concerns and practices noted in this report, they cannot be reported on in detail in advance of the designs themselves. However, future project deliverables (D4.2) will collate a number of the details of this process, in line with the timelines proposed for development of detailed designs.

# 5 REFERENCES

Ackerman, M S, Cranor, L F and Reagle, J (1999) Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences, *Proceedings of E-COMMERCE 99*, New York; ACM, 1-8

Adams A, Sasse MA (1999) Taming the wolf in sheep's clothing: privacy in multimedia communications. In: *Proceedings of the 7th ACM international conference on multimedia*, Orlando, Florida, October/November 1999, pp 101–107

Acquisti, A and Gross, R (2006) Imagined communities: Awareness, information sharing, and privacy on the Facebook, *Privacy Enhancing Technologies*, Springer

Al-Fedaghi, S (2007) How sensitive is your personal information?, *Proceedings of SAC'07*, ACM

Anderson, Jonathan, Claudia Diaz, Joseph Bonneau, and Frank Stajano. (2009) Privacy-enabling social networking over untrusted networks. In *Proceedings of the 2nd ACM workshop on Online social networks (WOSN '09)*. ACM, New York, NY, USA, 1-6. DOI=10.1145/1592665.1592667 http://doi.acm.org/10.1145/1592665.1592667

Anthony, D., D. Kotz, and T. Henderson. (2007) Privacy in location-aware computing environments. *IEEE Pervasive Computing,* 6(4):64–72, 2007.

Austerberry, D (2004) *Digital Asset Management*, Focal Press

Balebako, R, Jung, J, Lu, W, Cranor, L F and Nguyen, C (2013) "Little brothers watching you": raising awareness of data leaks on smartphones, *Proceedings of SOUPS '13*, New York: ACM

Bao, L and Intille, S S (2004) Activity recognition from user-annotated acceleration data, *Pervasive Computing*, Springer

Barbaro M, Zeller T Jr (2006) A Face Is Exposed for AOL Searcher No. 4417749. *New York Times*.

Barkhuus. L. (2004) Privacy in location-based services, concern vs. coolness. In *Proc. of the Workshop on Location System Privacy and Control,* 2004.

Barkhuus,L and A. Dey. (2003) Location-based services for mobile telephony: a study of users' privacy concerns. In *Proc. of INTERACT,* 2003.

Barnes, S B (2006) A privacy paradox: Social networking in the Unites States, *First Monday*, Volume 11, No 9, 4 September 2006

Barua, D, Kay, J Kummerfeld, B and Paris, C (2011) Theoretical foundations for user-controlled forgetting in scrutable long term user models, *Proceedings of OzCHI '11*, New York: ACM

Baumer, D. L. , J. B. Earp, and P. S. Evers, (2003) "Tit for tat in cyberspace: Consumer and website responses to anarchy in the market for personal information," *North Carolina Journal of Law and Technology,* vol. 4, no. 2, pp. 217–274.

Beach, S, Schulz, R, Downs, J, Matthews, J, Barron, B, and Seelman, K (2009) Disability, Age, and Informational Privacy Attitudes in Quality of Life Technology Applications: Results from a National Web Survey, *Transactions on Accessible Computing (TACCESS),* Volume 2 Issue 1, New York: ACM

Bellare, M., Boldyreva, A. & O'Neill, A., 2007. Deterministic and Efficiently Searchable Encryption. In *Advances in Cryptology - CRYPTO 2007*. Springer Berlin Heidelberg, pp. 535–552.

Bellotti, V and Sellen, A (1993) Design for privacy in ubiquitous computing environments, *Proceedings of ECSCW '93*, Springer

Ben-Asher, N., N. Kirschnick, H. Sieger, J. Meyer, A. Ben-Oved, and S. Möller. (2011) On the need for different security methods on mobile phones. In *Proc. of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI),* 2011.

Ben-Or, Michael, Shafi Goldwasser, and Avi Wigderson. (1988) Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *Proceedings of the twentieth annual ACM Symposium on Theory of computing, STOC '88*, pages 1–10. ACM.

Bendlin, R, I. Damgård, C. Orlandi, S. Zakarias. (2011) Semi-Homomorphic Encryption and Multiparty Computation. *EUROCRYPT 2011*.

Berendt, B, Gunther, O., and Spiekermann, S (2005) Privacy in E-Commerce: Stated Preferences vs. Actual Behavior, *Communications of the ACM*, April 2005, Vol. 48, No. 4, 101-106

Blasbalg, J., Cooney, R. and Fulton, S. (2012) Defining and Exposing Privacy Issues with Social Media, *Proceedings of JCSC 2012*, USAFA

Blass, E.O. et al., 2012. PRISM–privacy-preserving search in MapReduce. *Privacy Enhancing Technologies*. Available at: http://link.springer.com/chapter/10.1007/978-3-642-31680-7_10.

Boneh, D., G. DiCrescenzo, R. Ostrovsky, and G. Persiano. (2004) Public key encryption with keyword search. In *Proceedings of Eurocrypt 2004*. ISBN 978-3-540-72539-8. Barcelona, Spain: LNCS 3027, pp. 506-522.

Boneh, D., E. Kushilevitz, and R. Ostrovsky. (2007) Public Key Encryption that Allows PIR Queries. In *Proceedings of Crypto 2007*. ISBN 978-3-540-74142-8. Santa Barbara, USA: LNCS 4622, pp. 50–67.

Bonneau, Joseph, Jonathan Anderson, and George Danezis. (2009a) "Prying data out of a social network." Social Network Analysis and Mining. *ASONAM'09. International Conference on Advances in. IEEE.*

Bonneau, J., Anderson, J., Anderson, R., & Stajano, F. (2009). Eight friends are enough: social graph approximation via public listings. In *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems* (pp. 13-18). ACM.

Boyle, M and Greenberg, S (2005) The language of privacy: Learning from video media space analysis and design, *Transactions on Computer-Human Interaction (TOCHI),* Volume 12 Issue 2, New York: ACM

Boyles, J.L., A. Smith, and M. Madden. Privacy and data management on mobile devices. *Pew Internet and American Life Project*, Aug. 2012

Brodie, C, Karat, C-M., and Karat, J (2005) Usable Security and Privacy: A Case Study of Developing Privacy Management Tools, *Proceedings of Symposium on Usable Privacy and Security (SOUPS) 2005*, ACM

Brody et al (2012) *Social Networking and Impression Management: Self-Presentation in the Digital Age,* Lexington

Brush, A.J., Krumm, J. and Scott, J. (2010). Exploring End User Preferences for Location Obfuscation, Location-Based Services, and the Value of Location. In *Ubicomp '10*, 95-104.

Cachin, C., S. Micali, and M. Stadler. (1999) Computationally private information retrieval with polylogarithmic communication. In: *Proceedings of Advances in Cryptology, EUROCRYPT*. Vol. 1592. ISBN 3-540-65889-0. Prague, Czech Republic: LNCS, , pp. 402–414.

Cadiz, J. and Gupta, A. Privacy Interfaces for Collaboration. Microsoft Research. MSR-TR-2001-82 (2001).

Calandrino, J. A., Kilzer, A.,Narayanan, A., Felten, E.W, .and Shmatikov V. (2011) "You Might Also Like:" Privacy Risks of Collaboriate Filtering. In *IEEE Symposium on Security and Privacy*, IEEE Computer Society, 231-246, <https://www.cs.utexas.edu/~shmat/shmat_oak11ymal.pdf>.

Capra, R and Teevan, J (2012) Personal information management in a socially networked world, *Proceedings of CSCW'12*, New York: ACM

Castelluccia C. and A. Chan and E. Mykletun and G. Tsudik. (2009) Efficient and Provably Secure Aggregation of Encrypted Data in Wireless Sensor Networks. ACM ToSN (PDF) (*Transaction on Sensor Networks*).

Chakraborty, S, Raghavan, K R, Johnson, M P and Srivastava, M B (2013) A framework for context-aware privacy of sensor data on mobile systems, *Proceedings of HotMobile'13*, New York: ACM

Chang, Y-C, and M. Mitzenmacher. (2005) Privacy preserving keyword searches on remote encrypted data. In *Proceedings of Applied Cryptography and Network Security (ACNS)*. Vol. 3531, pp.442-455, ISBN 3-540-26223- 7. LNCS.

Chawla, Shuchi , Cynthia Dwork, Frank McSherry, Adam Smith, and Hoeteck Wee. (2005) "Toward Privacy in Public Databases." In: *TCC'05*, pp. 363–385.

Chellappa, R.K., (2002) Consumers' trust in electronic commerce transactions: The role of perceived privacy and perceived security, Retrieved from http://www.bus.emory.edu/ram/Papers/sec-privpdf 2002.

Chen, Y and Jones, G J F (2010) Augmenting human memory using personal lifelogs, *Proceedings of AH'10*, New York: ACM

Chen, Y and Xu, H (2013) Privacy management in dynamic groups: understanding information privacy in medical practices, *Proceedings of CSCW '13*, New York: ACM

Chin E, Felt A P, Sekar, V and Wagner, D (2012) Measuring User Confidence in Smartphone Security and Privacy, *Proceedings of Symposium on Usable Privacy and Security (SOUPS) 2012*, Washington, DC

Chong, S and Treiblmaler, H (2011) Trust and Perceived Risk of Personal Information as Antecedents of Online Information Disclosure: Results from Three Countries, *Journal of Global Information Management*, Volume 19, Issue 4

Chor, B., O. Goldreich, E. Kushilevitz, and M. Sudan. (1995) Private information retrieval. In: *Proceedings of the 36th Annual IEEE Symposium on Foundations of Computer Science*, pp. 41–51Milwaukee, USA.

Clarke, J, Castro, R R, Sharma, A, Lopez, J and Suri, N (2012) Trust & Security RTD in the internet of things: opportunities for international cooperation, *Proceedings of SecurIT'12*, New York: ACM

Consolvo, S., I. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge, (2005) "Location disclosure to social relations: Why, when, and what people want to share," in *Proceedings of CHI 2005, Conference on Human Factors in Com- puting Systems*, pp. 82–90, ACM Press, 2005.

Consolvo, S., J. Jung, B. Greenstein, P. Powledge, G. Maganis, and D. Avrahami. (2010) The Wi-Fi privacy ticker: improving awareness & control of personal information exposure on wi-fi. In *Proc. of Ubicomp*, 2010.

Conti, G. and Sobiesk, E. (2007) An Honest Man Has Nothing to Fear: User Perceptions on Web-based Information Disclosure, *Proceedings of Symposium on Usable Privacy and Security (SOUPS) 2007*, ACM

Conti, M, Hasani, A and Crispo, B (2013) Virtual private social networks and a facebook implementation, *Transactions on the Web (TWEB)*, Volume 7 Issue 3, New York: ACM

Crabtree, A, Mortier, R, Rodden, T and Tolmie, P (2012) Unremarkable Networking: The Home Network as a Part of Everyday Life, in *Proceedings of DIS 2012*, ACM

Crabtree, A, Rodden, T, Tolmie, P, Mortier, R, Lodge, T, Brundell, P, and Pantidi, N (2014) House Rules: The Nature and Role of Policy in Domestic Networks, *Personal and Ubiquitous Computing*, London: Springer-Verlag

Cramer, R., I. Damgaard, and J. Nielsen. (2001) Multiparty computation from threshold homomorphic encryption. *EUROCRYPT 2001.*

Cranor, L (2003) "'I didn't buy it for myself': Privacy and ecommerce personal- ization," in Proceedings of Workshop on Privacy in the Electronic Society, Washington, DC, USA: ACM Press

Cranor, L.F., Reagle, J., and Ackerman, M.S. (1999) *Beyond Concern: Understanding Net Users' Attitudes About Online Privacy*, AT&T Labs-Research Technical Report TR 99.4.3

Culnan, M. J. and P. K. Armstrong, (1999) "Information privacy concerns, procedu- ral fairness, and impersonal trust: An empirical investigation," *Organization Science*, vol. 10, no. 1, pp. 104–115.

Curtmola, R., J. Garay, S. Kamara, and R. Ostrovsky. (2006) Searchable symmetric encryption: improved definitions and efficient constructions. In *Proceedings of ACM Conference on Computer and Communications Security, CCS*. ISBN, pp.79-88. Alexandria, USA.

Cvrcek, D., M. Kumpost, V. Matyas, and G. Danezis. (2006) A study on the value of location privacy. In *Proc. of the 2006 Workshop on Privacy in an Electronic Society (WPES),* 2006.

Danezis, G., S. Lewis, and R. Anderson. (2005) How much is location privacy worth? In *Proceedings of the Workshop on the Economics of Information Security Series (WEIS),* 2005.

Denning, D.E. 1976. A lattice model of secure information flow. *Commun. ACM* 19, 5 (May 1976), 236-243. DOI=10.1145/360051.360056 http://doi.acm.org/10.1145/360051.360056

B. M. DePaulo and D. A. Kashy, (1998) "Everyday lies in close and casual relationships," *Journal of Personality and Social Psychology*, vol. 74, no. 1, pp. 63–79.

Dimakopoulos, D N and Magoulas G D (2009) Interface design and evaluation of a personal information space for mobile learners, *International Journal of Mobile Learning and Organisation*, Volume 3, Issue 4

Donath, J and Boyd, D (2004) Public Displays of Connection, *BT Technology Journal*, October 2004, Volume 22, Issue 4, 71-82

Doucet, L, Thatcher, S M B, and Thatcher, M E (2012) The effects of positive affect and personal information search on outcomes in call centres: An empirical study, *Decision Support Systems*, February 2012, Volume 52, Issue 3

Dourish, P., B. E. Grinter, J. D. D. L. Flor, and M. Joseph, (2004) "Security in the wild: User strategies for managing security as an everyday, practical problem," *Personal and Ubiquitous Computing*, vol. 8, no. 6, 2004.

Durrant, A, Golembewski, M, Kirk, D S, Beford, S, Rowland, D and McAuley, D (2011) Exploring a digital economy design space in theme parks, *Proceedings of DESIRE '11*, New York: ACM

Dwork, C (2006a) "Differential Privacy". In: *33rd International Colloquium on Automata, Languages, and Programming (ICALP).*

Dwork, C. (2006b) "Differential Privacy". In: *Automata, Languages and Programming*. Ed. by Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and IngoWegener. Vol. 4052. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2006, pp. 1–12. ISBN: 978-3-540-35907-4. URL: http://dx. doi.org/10.1007/11787006_1.

Dwork, C (2008a) "Differential privacy: a survey of results". In: *Proceedings of the 5th international conference on Theory and applications of models of computation. TAMC'08*. Xi'an, China: Springer- Verlag, pp. 1–19. ISBN: 3-540-79227-9, 978-3-540-79227-7. URL: http://dl.acm.org/ci tation.cfm?id=1791834. 1791836.

Dwork, C. (2008b) "Differential Privacy: A Survey of Results". In: *Theory and Applications of Models of Computation*. Vol. 4978. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 1–19.

Dwyer, C, Hiltz, S R, and Passerini, K (2007) Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace, *AMCIS 2007*

El Emam K, Jonker E, Arbuckle L, Malin B (2011) A Systematic Review of Re-Identification Attacks on Health Data. *PLoS ONE* 6(12): e28071. doi:10.1371/journal.pone.0028071

El Gamal. T. (1985) A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, 31(4):469–472.

Elkhiyaoui, K., Önen, M. & Molva, R., (2014) Privacy preserving delegated word search in the cloud. In *SECRYPT 2014, 11th International conference on Security and Cryptography, 28-30 August, 2014, Vienna, Austria*.

Ellison, N B, Steinfield, C, Lampe, C (2007) The benefits of Facebook "friends". Social capital and college students' use of online social network sites, *Journal of Computer-Mediated Communication*, Wiley

Emanuel, L, Bevan C and Hodges, D (2013) What does your profile really say about you?: privacy warning systems and self-disclosure in online social network spaces, Proceedings of CHI EA '13, New York: ACM

Fan, L. et al., 2000. Summary Cache: A Scalable Wide-area Web Cache Sharing Protocol. *IEEE/ACM Transactions on Networking*, 8(3), pp.281–293.

Felt, A.P., M. Finifter, E. Chin, S. Hanna, and D. Wagner. (2011) A Survey of Mobile Malware in the Wild. In *Proceedings of the ACM Workshop on Security and Privacy in Mobile Devices (SPSM),* 2011.

Felt, A., S. Egelman, M. Finifter, D. Akhawe, and D. Wagner. (2012a) How to ask for permission. In *Proc. of HotSec*, 2012.

Felt, A.P., Ha, E., Egelman, S., Haney, A., Chin, E., and Wagner, D., (2012b) Android Permissions: User Attention. Comprehension, and Behvior, *Proceedings of Symposium on Usable Privacy and Security (SOUPS) 2012*, ACM

Fischer, C. (1992). *America Calling: A Social History of the Telephone to 1940*. Berkeley: University of California Press.

Fox, S. Rainie, L., Horrigan, J., Lenhart, A., Spooner, T., and Carter, C. (2000). Trust and privacy online: Why Americans want to rewrite the rules. *The Pew Internet and American Life Project.* http://www.pewinternet.org/pdfs/PIP_Trust_Privacy_ Report.pdf

Friedman, B (1997) Social judgments and technological innovation: Adolescents' understanding of property, privacy, and electronic information, *Computers in Human Behavior*, Elsevier

Froehlich, J, Chen, M Y, Consolvo, S, Harrison, B, and Landay, J A (2007) MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones, Proceedings of MobiSys'07, New York: ACM

Furash, E. (1997). Leave me alone. *Journal of Lending & Credit Risk Man- agement*, 80, 62–65.

Gandy Jr, O H (1993) The Panoptic Sort: A Political Economy of Personal Information. *Critical Studies in Communication and in the Cultural Industries*, ERIC

Gemmell, J, Williams, L, Wood, K, Kueder, R, and Bell, G (2004) Passive capture and ensuing issues for a personal lifetime store, Proceedings of CARPE'04, New York: ACM, 48-55

Gennaro, R., Gentry, C. & Parno, B., 2010. Non-interactive Verifiable Computing: Outsourcing Computation to Untrusted Workers. In *Advances in Cryptology – CRYPTO 2010*. Springer Berlin Heidelberg, pp. 465–482.

Gentry, C., 2009. Fully homomorphic encryption using ideal lattices. *Stochastics. An International Journal of Probability and Stochastic Processes*. Available at: http://modular.math.washington.edu/home/wstein/www/home/watkins/CG.pdf.

Georgia Tech Graphics, Visualization & Usability Center (1998). *GVU's 10th WWW User Survey*. http://www.gvu.gatech.edu/user_surveys

Ghani, N A and Sidek, Z M (2008) Controlling your personal information disclosure, *Proceedings of ISP'08*, WSEAS

Ghani, N A and Sidek, Z M (2009) Personal information privacy protection in e-commerce, WSEAS *Transactions on Information Science and Applications*, March 2009, Volume 6, Issue 3

Gideon, J., S. Egelman, L. Cranor, and A. Acquisti. (2006) Power Strips, Prophylactics, and Privacy, Oh My! In *Proc. of the 2006 Symposium on Usable Privacy and Security,* pages 133–144, July 2006

Goffman, Erving (1959) *The presentation of self in everyday life*, New York: Anchor Books

Goldreich, O. and R. Ostrovsky. (1996) Software protection and simulation on oblivious ram. In: *Journal of the ACM* 45 (1996). ISSN 0004-5411, pp. 431–473.52. 1996

Good, N., R. Dhamija, J. Grossklags, S. Aronovitz, D. Thaw, D. Mulligan, and J. Konstan. (2005) Stopping spyware at the gate: A user study of privacy, notice and spyware. In *Proc. of the Symposium On Usable Privacy and Security (SOUPS),* 2005.

Good, N.S., and Krekelberg, A (2002) Usability and privacy: a study of Kazaa P2P file-sharing, HPL-2002-163, HP Labs

Greenberg, S and Rounding, M (2001) The notification collage: posting information to public and personal displays, *Proceedings of CHI'01*, New York: ACM, 514-521

Grinter, R. E. and L. Palen, (2002) "Instant messaging in teenage life," in *Proceedings of ACM Conference on Computer Supported Cooperative Work (CSCW2002*), pp. 21–30, ACM Press. http://doi.acm.org/10.1145/587078.587082, 2004.

Guha, Saikat, Bin Cheng, and Paul Francis (2011) Privad: practical privacy in online advertising. In *Proceedings of the 8th USENIX conference on Networked systems design and implementation (NSDI'11*). USENIX Association, Berkeley, CA, USA, 169-182.
<https://www.usenix.org/legacy/event/nsdi11/tech/full_papers/Guha.pdf>

Haddadi, H, Mortier, R, Hand, S, Brown, I, Yoneki, E, McAuley, D, Crowcroft, J (2012) Privacy analytics, *SIGCOMM Computer Communication Review*, Volume 42, Issue 3, New York: ACM

Hakkila, J. , and C. Chatfield, "(2005) Toward social mobility: 'It's like if you opened someone else's letter': User perceived privacy and social practices with SMS communication," in *Proceedings of Human Computer Interaction with Mobile Devices and Services MobileHCI '05,* pp. 219–222, Salzburg, Austria: ACM Press, http://doi.acm.org/10.1145/1085777.1085814, September 2005.

Hann, I-H., Hui, K.-L., Lee, T.S. & Png, I.P.L. (2002). Online information privacy: Measuring the cost-benefit tradeoff. *Proc. 23rd International Conference on Information Systems*.

Hardt, Michaela and Suman Nath. (2012) Privacy-aware personalization for mobile advertising. In *Proceedings of the 2012 ACM conference on Computer and communications security (CCS '12).* ACM, New York, NY, USA, 662-673. DOI=10.1145/2382196.2382266 http://doi.acm.org/10.1145/2382196.2382266

HAT (Hub of All Things) (2014) Platform for a Multi-Sided Market powered by the Internet of Things. hubofallthings.com

Hawkey, K and Inkpen, K M (2006) Keeping up appearances: understanding the dimensions of incidental information privacy, *Proceedings of CHI '06*, New York: ACM

Hoffman, D L, Novak, T P and Peralta, M (1999) Building consumer trust online, *Communications of the ACM*, 1999

Holone, H and Herstad, J (2010) Negotiating Privacy Boundaries in Social Applications for Accessibility Mapping, *Proceedings of NordiCHI 2010*, ACM

Hong, X, Nugent, C, Mulvenna, M, McClean, S, Scotney, B and Devlin, S (2009) Evidential fusion of sensor data for activity recognition in smart homes, *Pervasive and Mobile Computing*, Volume 5, Issue 3, June 2009, 236-252

Hood, Leroy, (2008) ""Systems biology and systems medicine: From reactive to predictive, personalized, preventive and participatory (P4) medicine"," Engineering in Medicine and Biology Society, 2008. EMBS 2008. *30th Annual International Conference of the IEEE*, vol., no., pp.cliv,cliv, 20-25 Aug. 2008. doi: 10.1109/IEMBS.2008.4649061.

Hood, L. & Stephen H. Friend. (2011) Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nature Reviews Clinical Oncology* 8, 184-187 (March 2011) |doi:10.1038/nrclinonc.2010.227

Hsieh, G., Tang, K.P., Low, W.Y., et al. (2007) Field Deployment of Imbuddy: A Study of Privacy Control and Feedback Mechanisms for Contextual Im. In *Ubicomp '07*, 91-108.

Hu, Hongxin, Gail-Joon Ahn and Jan Jorgensen. Detecting and Resolving Privacy Conflicts for Collaborative Data Sharing in Online Social Networks. In *Proceedings of 27th Annual Computer Security Applications Conference (ACSAC 2011),* Orlando, Florida, USA, December 5-9, 2011.

Huberman, B.A., E. Adar, and L. R. Fine. (2005) Valuating privacy. *IEEE Security and Privacy*, 3:22–25, 2005.

Iachello G and Hong, J (2007) End-User Privacy in Human-Computer Interaction, *Foundations and Trends in Human-Computer Interaction*, Vol. 1, No. 1 (2007) 1-137

Iachello, G., I. Smith, S. Consolvo, M. Chen, and G. D. Abowd. (2005) Developing privacy guidelines for social location disclosure applications and services. In *Proc. of the Symposium on Usable Privacy and Security (SOUPS),* 2005.

International Labor Organization (ILO) (1993) "Workers Privacy Part II: Monitoring and Surveillance in the Workplace Conditions of Work," Special Series on Workers Privacy, Digest 12, no. 1, 1993.

Ito, M and O. Daisuke, (2003) "Mobile phones, Japanese youth and the replacement of social contact," in *Front Stage/Back Stage: Mobile Communication and the Renegotiation of the Social Sphere*, (R. Ling and P. Pedersen, eds.), pp. 65–76, Grimstad, Norway, 2003.

Joinson, A N (2008) Looking at, looking up or keeping up with people?: motives and use of Facebook, *Proceedings of CHI'08*, New York: ACM

Jones, W and Anderson, K M (2011) Many views, many modes, many tools … one structure: Towards a Non-disruptive Integration of Personal Information, *Proceedings of HT'11*, ACM

Joye, M. & Libert, B., 2013. A Scalable Scheme for Privacy-Preserving Aggregation of Time-Series Data. In *Financial Cryptography and Data Security*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 111–125.

Jung, J., S. Han, and D. Wetherall. (2012) Short paper: Enhancing mobile application permissions with runtime feedback and constraints. In Proc. of the workshop on Security and Privacy in Smartphones and Mobile devices, 2012.

Kaasten, S., Greenberg, S. and Edwards, C. How People Recognize Previously Seen WWW Pages from Titles, URLs and Thumbnails. In *Proc. of Human Computer Interaction 2002*, Springer Verlag (2002), 247-265.

Kamara, S., Papamanthou, C. & Roeder, T., 2012. Dynamic Searchable Symmetric Encryption. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*. CCS '12. New York, NY, USA: ACM, pp. 965–976.

Kapadia, A, Henderson, T Fielding, J J, and Kotz, D (2007) Virtual walls: protecting digital privacy in pervasive environments, *Proceedings of PERVASIVE'07*, Springer-Verlag

Karkkainen, T, Vaittinen, T and Vaananen-Vainio-Mattila, K (2010) I don't mind being logged, but want to remain in control: a fied study of mobile activity and context logging, *Proceedings of CHI '10*, New York: ACM

Karr-Wisniewski, P., D. Wilson, and H. Richter-Lipford. (2011) A new social order: Mechanisms for social network site boundary regulation. In *AMCIS 2011 Proceedings, AMCIS '11*, page 9, Aug. 2011.

Kehoe, C., Pitkow, J., & Morton, K. (1997). Eighth WWW user survey. Retrieved from http://www.cc.gatech.edu/gvu/user_surveys/ survey-1997–04

Kelley, P.G., L. Cesca, J. Bresee, and L. F. Cranor. (2010)Standardizing privacy notices: an online study of the nutrition label approach. In *Proc. of CHI, 2010.*

Kelley, P., S. Consolvo, L. Cranor, J. Jung, N. Sadeh, and D. Wetherall. (2012) A conundrum of permissions: Installing applications on an android smartphone. In *Proc. of USEC*, 2012.

Kelly, L, Chen Y, Fuller, M and Jones, G J F (2008) A study of remembered context for information access from personal digital archives, *Proceedings of IIiX '08*, New York: ACM

Kim, P, Podlaseck, M and Pingali, G (2004) Personal chronicling tools for enhancing information archival and collaboration in enterprises, *Proceedings of CARPE'04*, New York: ACM

Kirk, D S and Sellen, A (2010) On human remains: Values and practice in the home archiving of cherished objects, *Transactions on Computer-Human Interaction (TOCHI),* Volume 17 Issue 3, New York: ACM

Kisilevich, S. and Mansmann, F. (2010) Analysis of Privacy in Online Social Networks of Runet, *Proceedings of SIN'10*, ACM

Klasnja, P., Consolvo, S., Jung, J., Greenstein, B.M., LeGrand, L., Powledge, P. and Wetherall, D. (2009) "When I am on Wi-Fi, I am Fearless:" Privacy Concerns & Practices in Everyday WiFi Use, *Proceedings of CHI 2009*, ACM

Kobsa, A., S. Patil, and B. Meyer. (2012) Privacy in instant messaging: An impression management model. *Behaviour & Information Technology*, 31(4):355–370, 2012.

KPMG International. (2011) The converged lifestlye. http://www.kpmg.com/convergence

Krause, A, Smailagic, A and Siewiorek, D P (2006) Context-Aware Mobile Computing: Learning Context-Dependent Personal Prefernces from a Wearable Sensor Array, *IEEE Transactions On Mobile Computing*, Vol. 5, No. 2, IEEE, 113-127

Krishnan, A and Jones, S (2005) TimeSpace: activity-based temporal visualisation of personal information spaces, *Personal and Ubiquitous Computing*, January 2005, Volume 9, Issue 1

Kwan, G C E and Skoric, M M (2013) Facebook bullying: An extension of battles in school, *Computers in Human Behavior*, Volume 29 Issue 1, Elsevier

Lahlou, S., Langheinrich, M., Röcker, C., (2005) Privacy and trust issues with invisible computers, *Communications of the ACM,* 48, 59-60, 2005.

Lampinen, A., S. Tamminen, and A. Oulasvirta. (2009) All my people right here, right now: management of group co-presence on a social networking site. In *Proceedings of the ACM 2009 international conference on Supporting group work, GROUP '09*, pages 281–290, New York, NY, USA, 2009. ACM.

Lawton, M.P., & Bader, J. (1970). Wish for privacy by young and old. *Jour- nal of Gerontology*, 25, 48–54.

Lederer, S., Dey, A. K., Mankoff, J. (2003). Who wants to know what when? Privacy preference determinants in ubiquitous computing *CY* 2003 Shortpapers.* 724-725.

Lederer, S, Hong, J I, Dey, A K and Landay, J A (2004) Personal privacy through understanding and action: five pitfalls for designers, *Personal and Ubiquitous Computing*, 2004, 8: 440-454

Lehtinen, V., Nasanen, J and Sarvas, R (2009) "A Little Silly and Empty-Headed" – Older Adults' Understandings of Social Networking Sites, *Proceedings of HCI 2009*, British Computer Society

Lendenmann. K.W. (2010) Consumer perspectives on online advertising. Technical report, Preference Central http://www.slideshare.net/mfredactie/ preference-central-surveyfullreport, 2010.

Lenhart, A and Madden, M (2007) Teens, privacy & online social networks: How teens manage their online identities and personal information in the age of MySpace, Pew Internet & American Life Project

Lessig, L (1999) The architecture of privacy, *Vanderbilt Journal of Entertainment Law & Practice*, Vol. 1, p. 56, 1999

Li, I, Forlizzi, J and Dey, A (2010) Know thyself: monitoring and reflecting on facets of one's life, Proceedings of CHI EA'10, New York: ACM

Lin, J, Amini, S, Hong, J I, Sadeh, N, Lindqvist, J and Zhang, J (2012) Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing, *Proceedings of UbiComp '12*, New York: ACM

Liu, Y., Gummadi, K.P., Krishnamurthy, B., and Mislove, A. (2011) Analyzing Facebook Privacy Settings: User Expectations vs. Reality, *Proceedings of IMC'11*, ACM

Livingstone, S (2008) Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression, *New Media & Society*, June 2008, vol. 10, no. 3, 393-411

Luger, E, Moran, S and Rodden, T (2013) Consent for All: Revealing the Hidden Complexity of Terms and Conditions, *Proceedings of CHI 2013*, New York: ACM, 2687-96

Luger, E and Rodden, T (2013) An Informed View on Consent for Ubicomp, *Proceedings of UbiComp'13*, New York: ACM, 529-38

Mancini, Clara, Yvonne Rogers, Keerthi Thomas, Adam N. Joinson, Blaine A. Price, Arosha K. Bandara, Lukasz Jedrzejczyk and Bashar Nuseibeh (2011) In the best families: tracking and relationships. *Proceedings of CHI '11*, ACM, New York.

Mao, H, Shuai, X, and Kapadia, A (2011) Loose Tweets: An Analysis of Privacy Leaks on Twitter, *Proceedings of WPES'11*, ACM

Marshall, N.J. (1974). Dimensions of privacy preferences. *Multivariate Behavioral Research*, 9, 255–271.

Martin, H, Bernados, A M, Iglesias, J and Casar, J R (2013) Activity logging using lightweight classification techniques in mobile devices, *Personal and Ubiquitous Computing*, Volume 17 Issue 4, Springer-Verlag

McAuley, D., Mortier, R., Goulding, J. (2011) The Dataware Manifesto. *Proceedings of the Third International Conference on Communication Systems and Networks (COMSNETS),* January 2011 .

McDonald, A.M., and L. F. Cranor. (2010) Beliefs and behaviors: Internet users' understanding of behavioral advertising. *TPRC 2010.*

McSherry, F.D. 2009. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data (SIGMOD '09),* Carsten Binnig and Benoit Dageville (Eds.). ACM, New York, NY, USA, 19-30. DOI=10.1145/1559845.1559850 http://doi.acm.org/10.1145/1559845.1559850

McSherry, F., and Mironov, I., (2009) Differentially private recommender systems: building privacy into the net. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09).* ACM, New York, NY, USA, 627-636.
<http://research.microsoft.com/pubs/80511/NetflixPrivacy.pdf>

Metzger, M. (2004). Privacy, trust, and disclosure: Exploring barriers to electronic commerce. *Journal of Computer-Mediated Communication*, 9(4). Retrieved January 15, 2006, from http://jcmc.indiana/edu/vol9/ issue4/metzger.html

Metzger, M., & Docter, S. (2003). Public opinion and policy initiatives for online privacy protection. *Journal of Broadcasting & Electronic Media*, 47(3), 350–374.

Milberg, S J, Burke, S J, Smith, H J and Kallman, E A (1995) Values, personal information privacy, and regulatory approaches, *Communications of the ACM*, Volume 38 Issue 12, Dec. 1995, 65-74

Milne, G.R., & Rohm, A.J. (2000). Consumer privacy and name removal across direct marketing channels: Exploring opt-in and opt-out alterna- tives. *Journal of Public Policy and Marketing*, 19, 238–249.

de Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific reports*, 3.

Morris, M E, Consolvo, S, Munson, S, Patrick, K, Tsai, J and Kramer, A D I (2011) Facebook for health: opportunities and challenges for driving behavior change, *Proceedings of CHI EA '11*, New York: ACM

Mortier, R et al (2010) The personal container, or your life in bits, *Proceedings of Digital Futures*

Mortier, R., H. Haddadi, T. Henderson, D. McAuley, J. Crowcroft (2013). Challenges & Opportunities in Human-Data Interaction. The Fourth Digital Economy All-hands Meeting: Open Digital (DE). November 4–6, 2013. Salford, UK

Multisilta, J and Milrad, M (2009) Sharing Experiences with Social Mobile Media, *Proceedings of MobileHCI'09,* New York: ACM

Mun, M, Hao, S, Mishra, N, Shilton, K, Burke, J, Estrin, D, Hansen, M, and Govindan, R (2010) Personal data vaults: a locus of control for personal data streams, *Proceedings of Co-NEXT '10*, New York: ACM

Murphy, M S (2011) Notes toward a politics of personalization, *Proceedings of iConference'11*, New York: ACM

Narayanan, Arvind, and Vitaly Shmatikov (2008) "Robust de-anonymization of large sparse datasets." *IEEE Symposium on Security and Privacy, 2008. SP 2008.*. IEEE.

Nath, S, Lin, F X, Ravindranath, L and Padhye, J (2013) SmartAds: bringing contextual ads to mobile apps, *Proceedings of MobiSys '13*, New York: ACM

New York Times (NYT) (2011) Got Twitter? You've Been Scored: http://www.nytimes. com/2011/06/26/sunday-review/26rosenbloom.html

Ng ICL (2013) *Value and Worth: Creating New Markets in the Digital Economy*. Innovorsa Press, Cambridge

Nguyen, D H and Hayes, G R (2010) Information privacy in institutional and end-user tracking and recording technologies, *Personal and Ubiquitous Computing*, Volume 14, Issue 1, Springer

Nikolaenko, V. et al., 2013. Privacy-Preserving Ridge Regression on Hundreds of Millions of Records. In *Security and Privacy (SP), 2013 IEEE Symposium on*. pp. 334–348.

Nissenbaum. H (2004) Privacy as contextual integrity. *Wash. L. Rev*., 79:119, 2004.

Ogata, W. and K. Kurosawa. (2004) Oblivious keyword search. *In vol. 20*. ISSN 0885-064X, pp. 356–371.

Oka, M, Hope, T, Hashimoto, Y, Uno, R and Lee, M-H (2011) A collective map to capture human behavior for the design of public spaces, *Proceedings of CHI EA'11*, New York: ACM

Oleksik, G and Brown, L M (2008) Sonic gems: exploring the potential of audio recording as a form of sentimental memory capture, *Proceedings of BCS-HCI'08*, British Computer Society

Olson, J.S., Grudin, J. and Horvitz, E. (2004) Toward Understanding Preferences for Sharing and Privacy

Olson, J.S., J.Grudin,and E.Horvitz,(2005) "A study of preferences for sharing and privacy," in *Proceedings of CHI '05 Extended Abstracts on Human Factors in Computing Systems*, April 2005.

O'Neil, D. (2001). Analysis of Internet users' level of online privacy concerns. *Social Science Computer Review*, 19, 17–31.

Önen, M. and R. Molva. (2007) Secure data aggregation with multiple encryption. *EWSN, 2007.*

Ongtang, M., S. McLaughlin, W. Enck, and P. McDaniel. (2009) Semantically rich application-centric security in Android. In *Proc. of the 25th Annual Computer Security Applications Conference (ACSAC),* December 2009.

Ostrovsky, R. and W.E. Skeith. (2007) A survey of single-database private information retrieval: techniques and applications. In: *Proceedings of the 10th international conference on Practice and theory in public-key cryptography*. ISBN 978-3-540-71676-1, pp. 393–411, Beijing, China.

P&AB Consumer Privacy Attitudes: A Major Shift Since 2000 and Why. *Privacy & American Business Newsletter* 10, 6 (2003), 1,3-5.

Paillier, P. (1999) Public-key cryptosystems based on composite degree residuosity classes. *EUROCRYPT 1999*.

Palen L, Dourish P (2003) Unpacking ''privacy'' for a networked world. In: *Proceedings of the CHI 2003 conference on human factors in computing systems*, Fort Lauderdale, Flor- ida, April 2003, pp 129–136

Panjwani, S., Shrivastava, N., Shukla, S. and Jaiswal, S. (2013) Understanding the Privacy-Personalization Dilemma for Web Search: A User Perspective, *Proceedings of CHI 2013*, ACM

Parke, R., & Sawin, D. (1979). Children's privacy in the home: Develop- mental, ecological and child-rearing determinants. *Environment and Behavior*, 11, 87–104.

Patil, S., and A.Kobsa,(2004) "Instant messaging and privacy,"in *Proceedings of HCI 2004*, pp. 85–88, Leeds, UK, 2004.

Patil, S. and J. Lai. (2005) Who gets to know what when: Configuring privacy permissions in an awareness application. In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '05*, pages 101–110, Portland, Oregon, USA, 2005. ACM.

Patil, S., Norcie, G, Kapadia, A and Lee, A.J. (2012) Reasons, Rewards, Regrets: Privacy Considerations in Location Sharing as an Interactive Practice, *Proceedings of the Symposium on Usable Privacy and Security (SOUPS) 2012*, ACM

Petersen, S. B. (1995) Your Life as an Open Book: Has Technology Rendered Personal Privacy Virtually Obsolete, *Federal Communications Law Journal*, 48: 163, 1995.

Phelps, J, Nowak, G and Ferrell, E (2000) Privacy Concerns and Consumer Willingness to Provide Personal Information, *Journal of Public Policy & Marketing*, Vol. 19, No.1, (Spring, 2000), 27-41

Popa, R.A., Catherine M. S. Redfield, Nickolai Zeldovich, and Hari Balakrishnan. 2011. CryptDB: protecting confidentiality with encrypted query processing. In *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles (SOSP '11).* ACM, New York, NY, USA, 85-100. DOI=10.1145/2043556.2043566 http://doi.acm.org/10.1145/2043556.2043566

Popa, R.A., Catherine M. S. Redfield, Nickolai Zeldovich, and Hari Balakrishnan. 2012. CryptDB: processing queries on an encrypted database. *Commun. ACM* 55, 9 (September 2012), 103-111. DOI=10.1145/2330667.2330691 http://doi.acm.org/10.1145/2330667.2330691

Prabaker, M, Rao, J., Fette, I., Kelley, P., Cranor, L., Hong, J. and Sadeh, N. (2007) Understanding and Capturing People's Privacy Policies in a People Finder Application, *Proceedings of UbiComp 2007*, ACM

Pratt, W, Unruh, K, Civan, A and Skeels, M M (2006) Personal health information management, *Communications of the ACM*, January 2006, Volume 49, Issue 1

Purcell, K., J. Brenner, and L. Rainie. (2012) Search engine use 2012. Technical report, March 2012.

Raij, A, Ghosh, A, Kumar, S and Srivastava, M (2011) Privacy risks emerging from the adoption of innocuous wearable sensors in the mobile environment, *Proceedings of CHI'11*, New York: ACM

Rawassizadeh, R, Tomitsch, M, Wac, K and Tjoa, A M (2013) UbiqLog: a generic mobile phone-based life-log framework, *Personal and Ubiquitous Computing*, Volume 17 Issue 4, Springer-Verlag

Sacks, H (1992) *Lectures on Conversation*, Blackwell

Sadeh, N. J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabaker, and J. Rao. (2009) Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing,* 13(6):401–412, 2009.

Sarin, S, Nagahashi, T, Miyosawa, T and Kameyama, W (2008) On the design and exploitation of user's personal and public information for semantic personal digital photograph annotation, *Advances in Multimedia*, Volume 2008, Issue 2, Hindawi

Schneier, Bruce, "A Taxonomy of Social Networking Data," *Security & Privacy, IEEE* , vol.8, no.4, pp.88, July-Aug. 2010. doi: 10.1109/MSP.2010.118. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5523874&isnumber=5523854

Schrammel, J, Koffel, C and Tscheligi, M (2009) How much do you tell?: information disclosure behaviour in different types of online communities, *Proceedings of C&T '09*, New York: ACM

Schwarz, PM (2003) Property, Privacy and Personal Data, *Harvard Law Review*

Sellen, A J, Fogg, A, Aitken, M, Hodges, S, Rother, C and Wood, K (2007) Do life-logging technologies support memory for the past?: an experimental study using sensecam, *Proceedings of CHI'07*, New York: ACM

Sellen, A J and Whittaker, S (2010) Beyond total capture: a constructive critique of lifelogging, *Communications of the ACM*, 2010, New York: ACM

Sheehan, K.B. (1999). An investigation of gender differences in on-line privacy concerns and resultant behaviors. *Journal of Interactive Market- ing,* 13(4), 24–38.

Sheridan, J, Bryan-Kinns, N, Reeves, S, Marshall, J, and Lane, G (2011) Graffito: crowd-based performative interaction at festivals, *Proceedings of CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*, ACM: New York, 1129-1134

Shi, E. et al., 2011. Privacy-Preserving Aggregation of Time-Series Data. *NDSS*. Available at: http://fxpal.com/publications/FXPAL-PR-11-636.pdf.

Shikfa, A., Önen, M. & Molva, R., 2011. Broker-Based Private Matching. In *Privacy Enhancing Technologies*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 264–284.

Sion, R. and B. Carbunar. (2007) On the Computational Practicality of Private Information Retrieval. In *Proceedings of Network and Distributed Systems Security Symposium (NDSS),* pp1-10. San Diego, USA.

Sleeper, M, Cranshaw, J, Kelley, P G, Ur, B, Acquisti, A, Cranor, L F and Sadeh, N (2013) "I read my Twitter the next morning and was astonished": a conversational perspective on Twitter regrets, *Proceedings of CHI'13, New York: ACM*

Smale, S and Greenberg, S (2006) Transient life: collecting and sharing personal information, *Proceedings of OZCHI'06,* ACM

Smith, H. J. , S. J. Milberg, and S. J. Burke, (1996) "Information privacy: Measur- ing individuals' concerns about organizational practices," *MIS Quart*, vol. 20, no. 2, pp. 167–196.

Sobradillo, P., Francisco Pozo, Álvar Agustí (2011) P4 Medicine: the Future Around the Corner, *Archivos de Bronconeumología ((English Edition)),* Volume 47, Issue 1, 2011, Pages 35-40, ISSN 1579-2129, http://dx.doi.org/10.1016/S1579-2129(11)70006-4. (http://www.sciencedirect.com/science/article/pii/S1579212911700064)

Song, D.X, D. Wagner, and A. Perrig. (2000) Practical Techniques for Searches on Encrypted Data. In *Proceedings of the IEEE Symposium on Security and Privacy*. pp. 44–55, Berkeley, California.

Song, J, Lee, S and Kim, J (2013) I know the shortened URLs you clicked on Twitter: inference attack using public click analytics and Twitter metadata, *Proceedings of WWW'13*, IWWWCSC

Spiekermann, S. (2005) "Perceived control: Scales for privacy in ubiquitous comput- ing," in *Proceedings of Conference on User Modeling — UM'05*, pp. 24–29, Edinburgh, UK, July 2005.

Spiekermann, S., Grossklags, J., and Berendt, B. (2001). E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior. *ACM Conference on Electronic Commerce (EC-01).* ACM Press, NY.

Stutzman F., and W. Hartzog. (2012) Boundary regulation in social media. In *Proceedings of ACM Conference on Computer Supported Cooperative Work, CSCW '12*, pages 769–778, Seattle, Washington, United States, 2012.

Sueda, K, Duh, H B-L and Rekimot, J (2012) Social like logging: can we describe our own personal experience by using collective intelligence?, *Proceedings of APCHI'12*, New York: ACM

Sweeney L (1997) Weaving technology and policy together to maintain confidentiality. *J Law Med Ethics* 25: 98–9110. doi: 10.1111/j.1748-720X.1997.tb01885.x.

Tang, K.P., Hong, J.I., Siewiorek, D.P. (2011) Understanding how visual representations of location feeds affect end-user privacy concerns, *Proceedings of UbiComp '11*, New York: ACM

Taylor, C., and Dajani, L. (2008) The future of homecare systems in the context of the ubiquitous web and its related mobile technologies, *Proceedings of PETRA'08*, New York: ACM

Taylor, H. (2003). Most people are 'privacy pragmatists' who, while concerned about privacy, will sometimes trade it off for other benefits. *Harris Interactive.* http://www.harrisinteractive.com/harris_poll/index.asp?PID=365.

Toch, E., J. Cranshaw, P. Hankes-Drielsma, J. Springfield, P. Kelley, L. Cranor, J. Hong, and N. Sadeh. (2010) Locaccino: A privacy-centric location sharing application. In *Proc. of the 12th ACM International Conference Adjunct Papers on Ubiquitous Computing,* 2010.

Tolmie, P. (2010) PA Work and the Management of Access to Information: Report of an ethnographic study of the work of Personal Assistants to inform the Personal Container theme within Horizon, Horizon Digital Economy Research

Tolmie, P. (2011) Uncovering the Unremarkable, in M.H. Szymanski and J. Whalen (eds) *Making Work Visible: Ethnographically Grounded Case Studies of Work Practice,* Cambridge University Press: Cambridge, 53-73

Tolmie, P., Crabtree, A. and Rouncefield, M. (2013) Mapping the Domestic Digital Economy: Findings from Ethnographic Studies of Domestic Settings – Report 1: An Overview of the Use of Digital Services in the Home, University of Nottingham: Horizon Digital Economy Research

Tolmie, P., Pycock, J., Diggins, T., MacLean A. and Karsenty, A. (2002) Unremarkable Computing. *Proceedings of CHI 2002, Conference on Human Factors in Computing Systems*, 2002, Minneapolis, Minnesota, 20-25 April 2002. New York: ACM, 399-406.

Toubiana, V., Narayanan, A., Boneh, D., Nissenbaum, H., & Barocas, S. (2010) Adnostic: Privacy Preserving Targeted Advertising. In *NDSS* (2010, February).
http://www.nyu.edu/pages/projects/nissenbaum/papers/adnostic.pdf

TRUSTe. 2011 consumer research results: Privacy and online behavioral advertising. http://www.truste.com/ad-privacy/TRUSTe- 2011- Consumer- Behavioral- Advertising-Survey-Results.pdf, July 2011.

Tsai, J., S. Egelman, L. Cranor, and A. Acquisti. (2007) The effect of online privacy information on purchasing behavior: An experimental study. In *Proc. of the Workshop on the Economics of Information Security,* 2007.

Turow, J., J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy. (2009) Americans reject tailored advertising and three activities that enable it. http://papers.ssrn.com/ sol3/papers.cfm?abstract_id=1478214, 2009.

UCLA Center for Communication Policy. (2000). The UCLA Internet re- port: Surveying the digital future: Year one. Retrieved October 7, 2001, from http://www.ccp.ucla.edu

UCLA Center for Communication Policy. (2001). The UCLA Internet report: Surveying the digital future: Year two. Retrieved January 21, 2002, from http://www.ccp.ucla.edu

UCLA Center for Communication Policy. (2003). The UCLA Internet re- port: Surveying the digital future: Year three. Retrieved October 1, 2005, from http://www.digitalcenter.org/pdf/InternetReportYearThree.pdf

UCLA Center for Communication Policy. (2004). The UCLA Internet re- port: Surveying the digital future: Year four. Retrieved October 1, 2005, from http://www.digitalcenter.org/downloads/DigitalFutureReport-Year4-2004.pdf

User-Centric Networking (UCN) (2013) FP7-Information and Communication Technologies: Call 10 Project Proposal

User-Centric Networking (UCN) (2014) D 5.1: Preliminary UCN Use Cases and Applications, EC FP7-ICT-2013-10, 1.6 Connected and Social Media, Grant Agreement No.: 611001

Ur, B, Leon, P D, Cranor, L F, Shay, R and Wang, Y (2012) Smart, Useful, Scary, Creepy: Perceptions of Online Behvioral Advertising, *Proceedings of Symposium on Usable Privacy and Security (SOUPS) 2012*, Washington, DC

Urban, J., C. Hoofnagle, and S. Li. (2012) Mobile phones and privacy. UC Berkeley Public Law Research Paper, 2012.

Vallet, D., Friedman, A., Berkovsky, S., (2014) Matrix Factorization Without User Data Retention. In *PAKDD 2014*, 569-580.<http://www.nicta.com.au/pub?doc=7467&filename=nicta_publication_7467.pdf>

Viegas, F.B. (2005) Bloggers' expectations of privacy and accountability: An initial survey, *Journal of Computer-Mediated Communication*, Wiley

Vijayan J., (2012) States Challenge Google Privacy Policy Change, *Computer World,* Retrieved from http://www.pcworld.com/article/250698/states_challenge_google_privacy_polic y_change.html, 2012.

Vincent, A., Kaelber, D., Pan, E., Shah, S., Johnston, D., Middleton, B. (2008). A Patient-Centric Taxonomy for Personal Health Records (PHRs). *AMIA Annu Symp Proc. 2008*; 2008: 763–767. Published online 2008.

Viswanath, B., Emre Kiciman, and Stefan Saroiu. 2012. Keeping information safe from social networking apps. In *Proceedings of the 2012 ACM workshop on Workshop on online social networks (WOSN '12)*. ACM, New York, NY, USA, 49-54. DOI=10.1145/2342549.2342561 http://doi.acm.org/10.1145/2342549.2342561

Wang, Y, Norcie, G, Komanduri, S, Acquisti, A, Leon, P G, and Cranor, L F (2011) "I regretted the minute I pressed share": a qualitative study of regrets on Facebook, Proceedings of SOUPS'11, New York: ACM

Warren, and Brandeis. (1890). The Right to Privacy. *Harvard Law Review, IV(5).*

Waters, B.R. et al., 2004. Building an Encrypted and Searchable Audit Log. *NDSS*. Available at: http://www.isoc.org/isoc/conferences/ndss/04/proceedings/Papers/Waters.pdf.

Watson, J. Besmer, A, Richter Lipford, H (2012) +Your Circles: Sharing Behavior on Google+, *Proceedings of Symposium on Usable Privacy and Security (SOUPS 2012*, ACM

Westin A (1967) *Privacy and freedom*. Atheneum, New York

Westin, Alan F. (1991). *Harris-Equifax Consumer Privacy Survey 1991*. Atlanta, GA: Equifax Inc.

Westin, Alan F. (1998). *E-commerce & Privacy: What Net Users Want*. Hackensack, NJ: Privacy & American Business.

Whitten A. and Tygar, J. D.. (1999) Why Johnny can't encrypt. In *Proceedings of the 8th USENIX Security Symposium*, August 1999.

Wiese, J., P. G. Kelley, L. F. Cranor, L. Dabbish, J. I. Hong, and J. Zimmerman. (2011) Are you close with me? Are you nearby?: Investigating social groups, closeness, and willingness to share. In *Proc. of the 13th International Conference on Ubiquitous Computing,* 2011.

Winckler, M, Gaits, V, Vo, D-B, Sergio, F and Rossi, G (2011) An approach and tool support for assisting users to fill-in web forms with personal information, *Proceedings of SIGDOC'11*, ACM

Wisniewski, P., H. Lipford, and D. Wilson. (2012) Fighting for my space: Coping mechanisms for SNS boundary regulation. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, CHI '12*, pages 609–618, New York, NY, USA, 2012. ACM.

Wogalter. M.S. (2006) Communication-Human Information Processing (C-HIP) Model. In *Handbook of Warnings.* Lawrence Erlbaum Associates.

Wolfe, M., & Laufer, R. (1974). The concept of privacy in children and adolescence. In D. Carson (Ed.), *Man–environment interactions: Evalua- tions and applications: Privacy* (Part 2, Vol. 6, pp. 29–54). Washington, DC: Environmental Design Research Association.

Yao, Andrew C. (1982) Protocols for secure computations. *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science*, pp 160–164.

Yao, M.Z., Rice, R.E. and Wallis, K. (2007) Predicting User Concerns About Online Privacy, *Journal of the American Society for Information Science and Technology*, 58(5): 710-722

Zhao, S, Grasmuck, S and Martin, J (2008) Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in human behavior*, Elsevier

Zhong, M, Liu, M and He, Y (2013) 3SEPIAS: A Semi-Structured Search Engine for Personal Information in Dataspace System, *Information Sciences: an International Journal*, Volume 218, Elsevier

Zhou, Y., Z. Wang, W. Zhou, and X. Jiang. (2012) Hey, You, Get Off of My Market: Detecting Malicious Apps in Official and Alternative Android Markets. In *Proceedings of the Network and Distributed System Security Symposium (NDSS),* 2012.

# APPENDIX

| Use Case Family | Use Case ID | Use Case Description | Third Party Access |
|---|---|---|---|
| **Recommender systems** | FF1 | UCN recommender system wants to segment all users' preferences into meaningful clusters | ✓ |
| **Recommender systems** | FF2 | UCN recommender system wants to be able assign a user to a cluster | ✓ |
| **Recommender systems** | FF3 | User should be allowed to keep a local "have seen list" that is not accessed by UCN's recommender system | ✘ |
| **Recommender systems** | FF4 | User should be allowed to remove irrelevant content from UCN's recommender system | ✓ |
| **Recommender systems** | FF5 | UCN's recommender system should be responsive to any change in users' interest | ✓ |
| **Smart homes** | IM01 | The PIH should adjust video streaming and upload to available bandwidth | ✘ |
| **Smart homes** | IM02 | The PIH should be able to identify and authenticate the user | ✘ |
| **Smart homes** | IM03 | Privacy preserving access to camera feeds that are outsourced to the cloud | ✘ |
| **Smart homes** | IM04 | Easy installation of security cameras | ✘ |
| **Smart homes** | IM05 | Storage at the PIH should be expendable | ✘ |
| **Smart homes** | IM06 | PIH should be able to stream video with trick play | ✘ |
| **Smart homes** | IM07 | PIH should be able to switch off power for individual plugs | ✘ |
| **Smart homes** | IM08 | PIH should be able to determine current energy flow | ✘ |
| **Smart homes** | IM09 | PIH should be able to read sensor data | ✘ |
| **Smart homes** | IM10 | PIH should be able to detect the presence or absence of individuals | ✘ |
| **Smart homes** | IM11<br>IM12 | User should be able to configure the gateway | ✘ |

| | | | |
|---|---|---|---|
| | IM13<br>IM14 | | |
| **Smart homes** | IM15 | The PIH should be able to notify the user when needed | ✘ |
| **Smart homes** | IM16 | The PIH should be able to switch between video storage options (locals Vs. cloud) | ✘ |
| **Smart homes** | IM17 | The PIH should be endowed with uninterruptible power supply | ✘ |
| **Smart homes** | TC01<br>TC02 | The PIH should be able to identify user based on visual cues | ✘ |
| **Smart homes** | TC03 | The PIH should be able to transfer AV information | ✘ |
| **Smart homes** | TC04 | The PIH should be able to determine parental rating | ✘ |
| **Smart homes** | TC05 | Based on external information, PIH should be able to adjust notifications for future events | ✘ |
| **Smart homes** | TC06 | The PIH should be able to determine whether the house in a safe state or not | ✘ |
| **Recommender systems** | PT01<br>PT02 | The PIH should show the user his most watched content | ✘ |
| **Recommender systems** | PT03<br>PT05 | The UCN recommender system should recommend content to a user based on his friends' interests | ✓ |
| **Smart homes** | PT04<br>PT07<br>PT08 | The user should be able to share content with his friends | ✓ |
| **Recommender systems** | PT06 | Based on the DVR content that the user is watching, the UCN recommender system adjust its advertisements | ✓ |
| **Smart homes**<br>**Recommender systems** | CO1 | The user should be able to control what information is given to third parties | ✘ |
| | CO2 | The user should be able to access the information stored at the PIH | ✘ |