Personal Data Management:

The User's Perspective

Note: This report is based primarily on research commissioned by the International Institute of Communications from IPSOS UU between March-August 2012, funded by the Microsoft Corporation.

Table of Contents

Preface 3

Executive Summary 7

The Findings 12

1 Methodological Considerations 12

1.1 Setting a Baseline: Users, Technology, and Awareness 13

2 Data Context 15

2.1 Type of Data 16

2.2 Type of Entity 17

2.3 Trust in the Service Provider 18

2.4 Collection Method 20

2.5 Device Context 23

2.6 Data Usage 23

2.7 Value Exchange 24

3 Understanding Risk 26

4 Perceptions of Accountability 29

4.1 Expectations of Self-Accountability 29

4.2 Accountability of Service Providers 30

4.3 Accountability of Government and Intermediaries 31

5 Expected Data Principles 33

6 Key Differences by Market 34

7 Conclusion 37

7.1 User Mental Models 37

7.2 What Does This Mean for Regulation? 39

APPENDIX A – Methodology 44

APPENDIX B – Detailed Research Survey Results 45

Personal Data Management: The User's Perspective

Preface

The International Institute of Communications (IIC) spans the communications industry, from broadcasting (where it began in 1969) to other forms of entertainment and business content, from fixed-line telecommunications to wireless services, from telephone handsets that required a table on which to 'sit' to smartphones, from data and video delivered using Internet Protocol-based transmission systems to data held in the cloud. With these changes have come an evolution in how individuals act and react with the different technologies they encounter. For example, the passive viewer in the analogue broadcasting world has not only become an active viewer in the digital broadcasting world, but also could be a producer and aggregator of their own content. This is compounded by a growing number of organizations that are collecting massive amounts of data pertaining to both passive and active viewers and participants in an endless range of activities.

It is clear from the activities that the IIC undertakes, in its meetings and through its journal, that these changes, which started as discrete technological developments, now overlap or parallel each other. And each change in how and what data are used incrementally increases the importance of how data is transmitted, exchanged and managed. Data and data flows have become vital to the development and growth of national economies. They can also bring significant societal benefits, both to individuals and to their communities and nations.

However these different elements, all part of a whole, are often treated as separate components by those who create policy and the IIC seeks to re-unite these various strands to form a holistic picture.

The dilemma of stakeholders, who must grapple with the volume of data being generated, transmitted and received is understandable. As the opening section of a recent report by The Pew Research Centre observes[1]:

"We swim in a sea of data … and the sea level is rising rapidly.

[1] http://pewinternet.org/Reports/2012/Future-of-Big-Data.aspx

"Tens of millions of connected people, billions of sensors, trillions of transactions now work to create unimaginable amounts of information. An equivalent amount of data is generated by people simply going about their lives, creating what the McKinsey Global Institute calls "digital exhaust" - data given off as a by-product of other activities such as their Internet browsing and searching or moving around with their smartphone in their pocket.

"Human-created information is only part of the story, a relatively shrinking part. Machines and implanted sensors in oceans, in the soil, in pallets of products, in gambling casino chips, in pet collars, and countless other places are generating data and sharing it directly with data "readers" and other machines that do not involve human intervention…..

"While enthusiasts see great potential for using Big Data, privacy advocates are worried as more and more data is collected about people - both as they knowingly disclose things in such things as their postings through social media and as they unknowingly share digital details about themselves as they march through life. Not only do the advocates worry about profiling, they also worry that those who crunch Big Data with algorithms might draw the wrong conclusions about who someone is, how she might behave in the future, and how to apply the correlations that will emerge in the data analysis."

As this extract suggests, and as much current policy indicates, a large part of the debate around data collection and data aggregation relates to the way in which data might be managed and how it is used; in particular, data which might be considered "personal". The term "personal data" is defined by the European Union as "*any information relating to an identified or identifiable natural person ('Data Subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*"[2]. With the exponential growth in data collection and traffic, there are concerns that people should be protected from harm that could be caused by the misuse of their data. However this has led, in many countries, to a regulatory approach which is arguably over-protective so that all data collected are defined as "personal" or "not personal" regardless of context, and regulation is applied depending on whether or not the "personal" bar is reached.

Such an approach is not sustainable in the emerging market for data usage. So, regulators are considering how best to create appropriate regulation without limiting or reducing the benefits

[2] EU Data Protection Directive (95/46/EC) - http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML

conferred on individuals and society at large by the digital economy. Creating such a balanced framework requires an understanding of the potential economic impact of the flow of data, the emerging technologies that can provide possible solutions, as well as individual attitudes and behaviours towards personal data.

This report focuses on the last aspect, and draws on original qualitative research, conducted by IPSOS UU. The objective is to develop evidence-based insights into individual *"mental models"* regarding personal data: i.e. what factors impact individuals' sensitivity to the collection, access, and use of their personal data, what are the perceived risks, and how willing are they to manage their data. These insights are integral to the development of a framework of regulation which answers the concerns of, and offers protection to, users in meaningful ways while allowing for the evolution of technologies and the societal and economic benefits it can bring.

The research is based on interviews with individuals and focus groups in four countries that have considerable variation in their approaches to regulating personal data, as well as differing social and cultural environments: Canada, China, Germany and the US. Qualitative research allows for a concentrated examination of issues and a rehearsal of possible scenarios, used to great effect in this study. As part of the interview methodology, participants had to consider different situations in which their personal data might be used[3].

The findings of this study show that the benefits of the digital services on offer outweigh significantly any perception of the potential for harm, or even concerns about possible risks. Many of the participants interviewed (all of whom are users of online technologies, and are referred to throughout the report as user-participants) accept a level of both personal accountability and personal responsibility for what they put online. Indeed most of them adopt a range of strategies to limit the sharing of sensitive data, such as creating multiple identities for different uses or inputting false data. Nonetheless the study also indicates user-participants' acceptance of the fact that control is largely relinquished once personal data are put online. It provides a view, from the user's perspective, of the actions stakeholders can take to mitigate their concerns about having to place so much trust in the services to which they have given data.

"*You have so much information that is shared. When it's out there, it's out there."* (Denver)

[3] See Appendix A.

The majority of this report presents the findings from this commissioned research and suggests how policymakers might address the issues raised by the user-participants.

Executive Summary

Aims and Scope of the Research

The term "personal data" encompasses a wide range of data about a person who can be identified, "directly or indirectly"[4]. In the European Union, citizens have a fundamental right to privacy (Article 8 of the Charter of Fundamental Rights of the EU)[5], and the right to be informed about whether and how data about them are collected, processed and transferred, including in the workplace[6]. This model implies that, in giving consent, the user is aware of the use to which their information is put.

The IIC commissioned a study from IPSOS UU (funded by Microsoft) to gain insights into the mental models of individuals on the collection, access and use of their data. The resulting report offers suggestions for regulatory frameworks that would more accurately reflect the nuances of these user mental models (that is, the sensitivities attached to personal information given to an internet service provider) in a world where personal data have become increasingly commoditised. The qualitative study used a mixture of focus groups and individual interviews with people who use online services[7]. It covers four markets: Canada (Toronto), China (Shanghai), Germany (Hamburg) and the US (Denver). These were chosen to reflect differing regulatory approaches to personal data management and differing social and cultural norms. Insights into the impact of these differences are especially important in the digital environment, where national boundaries are porous.

---

[4] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML

[5] http://www.europarl.europa.eu/charter/pdf/text_en.pdf

[6] From the Information Commissioner's Office, UK:
Personal data means data which relate to a living individual who can be identified –
(a) from those data, or
(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,
And includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.
It is important to note that, where the ability to identify an individual depends partly on the data held and partly on other information (not necessarily data), the data held will still be "personal data".

[7] See Appendix A.

The study was informed by other research such as that conducted by the Pew Research Centre and the Eurobarometer survey on attitudes to data protection in the European Union[8] and examined the following:

- What components are required in a more holistic policy framework that will enable and empower users to manage the access, flow and use of their personal data?
- Is there a balance to be struck between protecting users from harm and ensuring that data can be used to deliver societal/social and economic good? What are users willing to do to impact this balance?
- Is there some commonality in users' perception of identity, personal data rights, data access, etc., that can be leveraged to develop an interoperable technology and policy framework?
- What are users' attitudes to their own accountability and that of other stakeholders within the data ecosystem?

In particular the study sought to:

- Explore users' boundaries regarding the application and integration of their personal data when using the Internet, specifically:
    - The types of personal data collected;
    - The means of collecting personal data (i.e., active vs. passive[9]); and
    - The ways of using personal data.
- Understand the risks/harms that users associate with having their personal data collected and used.
- Gauge user attitudes to the value they attribute to providing their personal data.
- Identify which entities users trust with regard to the collection, use, and protection of their personal data (e.g., service providers, government, intermediaries).
- Understand motivations for user self-management of their online privacy and identity.
- Assess user expectations of their rights associated with the collection, use and protection of their personal data.

As the comment from the Pew Research above indicates, the use to which personal data are put varies and can be reconfigured in many ways based on the user's mental model and the context in

[8] http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf

[9] *Active data collection* results from a transaction where the user is an active party, inputting information as, for example, in an online purchase. *Passive data collection* occurs when the individual is unaware the transaction is taking place, such as in CCTV footage captured at a train station or location data captured while carrying out a search on a mobile device.

which personal data is being exchanged. Much current regulation considers the individual and how his or her personal data is handled; such regulation tends to be in the province of data protection/information/privacy regulators, either independent bodies or those that are part of another agency. Rules and guidelines often concentrate on protecting the individual from "the abuse or disclosure of intimate personal data"[10].

There are also (often separate) regulatory agencies which deal with regulation of data collected generically which are then aggregated. Some of these aggregated data can be disaggregated – and so there is regulation which concentrates on data storage and security. Some data are aggregated and remain so, but are used in ways for which the individual might not have given consent if he or she had understood such usage. There is then, a balance to be achieved between safeguarding the user's personal data and developing regulation that is so stringent that innovation is curbed because service providers are uncertain about the parameters within which they can develop new services; services that ultimately could benefit the economy and the user.

The opinion that results from these findings is not prescriptive but suggests some general principles that might be accommodated within a policy framework. The recommendations focus on regulatory policy; recommendations for technological developments exist to the extent that parameters that define user mental models are suggested, but there is no discussion of how these can or should be implemented.

Key Recommendations

The research suggests that a simplistic, binary personal data management policy is neither flexible nor appropriate in the fast-developing online environment. Instead, personal data usage is at best nuanced in the mind of the consumer, and user-participants generally understand that they accept a level of plausible risk when they sign up for a service; this, in turn, encourages them to put some control mechanisms in place.

Participants show that, depending on the data context, they apply different values (emotional and monetary) to the types of data they provide, and anticipate different responses in return. These expectations of responses are, in turn, based on assumed prospects of benefits, or indeed disadvantages. The insights from this study therefore, suggest there is a need for a holistic and nuanced policy framework of personal data management that:

[10] http://www.oecd.org/document/18/0,3746,en_2649_34223_1815186_1_1_1_1,00.html

- Is based on responsible and trusted data exchanges amongst stakeholders;
- Respects a set of personal data principles that empowers and protects users from harm;
- Recognises that users currently take actions to manage and control the data they share;
- Acknowledges that different levels of data sensitivity exist and there are multiple elements within the data context that define users' willingness to share personal data.

The interpretation of the study's findings defines six broad insights about users' mental models that policymakers should consider in developing a framework or guidelines. These are:

- **Trust and accountability**: While trust is a vital element in the data context, it is neither a necessary nor sufficient variable in the decision to share personal data. However, it is a variable of some importance. Therefore, all stakeholders involved in the personal data ecosystem (users, service providers, regulators, etc.) must accept their responsibility to provide a trusted environment within which users of the internet can be confident about the exchange and use of their data. They should demonstrate that they have processes of accountability in place to ensure compliance with existing regulatory structures, and redress in the event of breaches in data use policies.
- **Principles-based regulation:** To accommodate national interests and differences, and to allow flexibility as the market develops, stakeholders should develop high-level principles to create a framework that allows for certain key tenets to be followed, implemented, and enforced, thus allowing the user to develop and maintain trust in the digital environment.
- **Data context:** The study finds a number of variables that affect how user-participants think about their personal data. These include the type of data being volunteered, the service into which it is being transferred and how much they trust that service – all the user-participants in this study use different personas for different services, thereby adopting a self-regulatory approach to their personal data management[11]. Other important contextual variables are the use to which their information is put, whether or not they wish to give permission for such use, and what they feel they are getting in return. The type of device on which the transaction is being made also affects attitudes to personal data collection. These many variables argue for the nuanced regulatory approach to personal data management recommended here.
- **Graduated control**: The user-participants in this study suggest a graduated process of permissions, based on the context of data use, which includes the type and quantity of data being provided, the use to which they are to be put, and the relevance of any service being offered. These permissions, to be sought from the user, should be delivered in a way that

[11] By "personas", user-participants mean that they use different identities, as represented by the names or pseudonyms they give themselves, to express their online identity, or to keep their real identity masked.

ensures that the user takes more personal responsibility, basing his or her decisions on information received in return for the data given. This type of control differs from the binary on/off personal data switch that currently forms the basis of much regulation. Ideally, the default settings should take into consideration existing social and cultural norms as well as past user preferences.

- **Value exchange**: User-participants understand that, in order to access certain services (especially if they are free), they may need to give the service provider certain information. But they want to be assured they will receive benefit of comparable value in exchange for these data. User-participants suggest they may be willing to provide increasing amounts of data when they can see the benefits afforded to them by the service that uses those data.
- **Governance**: User-participants want service providers to be accountable to them but understand that this might be through third parties. It was suggested by some participants that an intermediary organisation might be appropriate, which suggests a co-regulatory structure might be viable with relevant compliance and enforcement powers. In all cases, user-participants expect that governments will have ultimate authority to determine regulatory action and enforce compliance.

The Findings

1 Methodological Considerations

This report is based on the findings from a qualitative study of internet users to establish their "mental models" on personal data management. It uses both depth interviews with individuals and group discussions in Canada, China, Germany, and the US. The four territories were chosen on the basis of their different data protection regulatory regimes and cultures. User-participants ranged in age from 21 to 60, with an even mix of male and female, single and married, and included full time students and a variety of professions. They were recruited against a battery of questions which tested their personalities on the following four profiles to provide a range of attitudes and behaviours of active internet users: technology enthusiasts who must always have the latest technology; technology users who are constantly on the move; social networkers who need to stay connected with friends; and technology pragmatists who use technology to manage their life and home activities. These profiles will be recruited against in the follow-up quantitative study. Further details of the sample and methodology are included in the Appendix A.

While this is a study on user behaviour and attitudes regarding the management of their data, it is worth noting that technology is evolving towards the increasing use of sensors and wireless devices which will collect data automatically, through machine-to-machine transactions. This means that users will not be involved with or even aware of the transactions taking place[12]. Such transactions might include location data from users' mobile phones as they search for information; video captures of users entering a train station; a thermostat in the home transferring data to a utility company so energy usage and prices can be determined in real time; or a car transferring data about current speed and road traction to a central facility to capture current traffic and road conditions. In the study, and in this report, these data are referred to as being *"passively collected data."* This is to contrast them with *"actively collected data,"* where the user is proactively volunteering information and/or participating in the transaction, e.g., the posting of a photo or a blog, entering a credit card number for an online purchase, or entering bank account details to enable an online banking transaction.

[12] By the end of 2012, the number of mobile connected devices will exceed the number of people on earth. By 2020, there will be 50B devices connected wirelessly to the Internet. Between 2012-2016, machine to machine traffic will grow 22 times to 5x1017/month (CAGR of 86%). ("Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011–2016," February 14, 2012.)

Another type of *"passively collected data"* is *"inferred data",* insights resulting from analytics that aggregate data collected about the user with other data sets, including data about past user transactions and transactions involving similar users.

The research technique specifically distinguished between actively collected data and the different types of passively collected data.

1.1 Setting a Baseline: Users, Technology, and Awareness

To establish a baseline for user-participant attitudes, all sessions started with a general discussion on users' perceptions of technology. This provides the context to calibrate and interpret the remainder of the discussion.

It is found that, overall, technology enables users to do more, either by making what they were doing more efficient and/or enabling them to do things that were not previously possible. As a result, user-participants generally have positive attitudes towards digital technologies due to the benefits brought to their personal and professional lives. None can imagine life without their devices and internet access.

"It's super that there's technology. There are many new things that make things easier in both my professional and private life." (Hamburg)

However, this also means that users become increasingly dependent on technology which they do not fully understand and this can make them feel powerless.

"*When it works, it works great. But everything is now dependent on technology." (Toronto)*

This feeling of powerlessness leads to a perceived need for greater control. Furthermore, since user-participants can be overwhelmed by technology, their perception of the level of trust they can have in a service or service provider can impact on their sense of ease in interacting with those online services or service providers. Thus, even at this early stage in the research, when the discussion is about technology in general and its benefits and detriments, prior to any discussion of personal data, it is clear that control and trust are critical elements in user mental models.

The research finds that user-participants may be characterized by four attributes as shown in Figure 1 below, but the potency of those attributes varies by country.

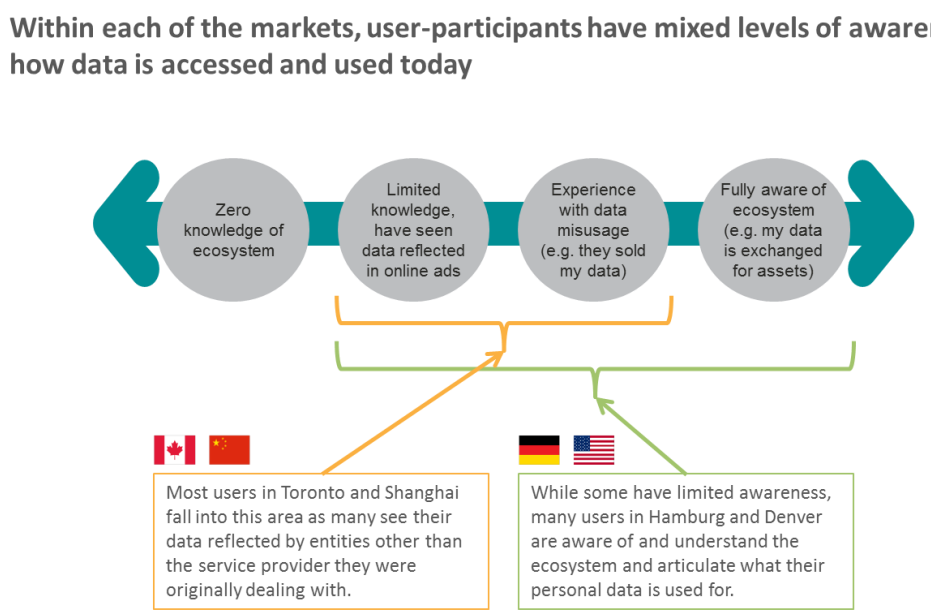
Figure 1: User type characterisation by country

Four personal attributes uncovered in this research make up the dimensions of user types, including:

Personal data awareness	Low	Low	Mid	Mid
Trust in government	High	Low	High	Low
Perceptions of own accountability	High	Low	High	High
Desire for control	High	High	High	High

"*Personal data awareness*" refers to how aware user-participants are about the data ecosystem into which they volunteer their personal data. In general, the research finds limited awareness, and little understanding, of the distinctions between first party (the application or service to which the user has provided his or her information) and third party (subsequent users of the information, including collected and inferred) users of personal data; see Figure 2.

Figure 2: Awareness of the personal data ecosystem



"Trust in government" reflects a potential correlation between a country's data protection regulatory framework and the user-participants' sense of trust in their government. Canada and Germany both have rigorous and relatively well-publicised approaches towards data protection, resulting in the somewhat high level of trust in government suggested by participants from those countries. The lack of data protection regulation in China and the more market-oriented approach in the US lead to a comparatively low level of trust in government reported in those countries.

In general, *"perceptions of own accountability"* was high as user-participants accept that they should be accountable for the personal information they put online, although they feel they are not responsible for any subsequent distribution of this information for other uses. User-participants' sense of accountability is discussed in more detail in Section 4 below. China is an exception to this. Here, user-participants express a sense of resignation regarding overall control of their lives as they are used to a more authoritarian regime, and their perceived lack of personal accountability seems to be exacerbated by the high percentage of accounts that are breached, resulting in the unauthorized modification of account contents.

The "*desire for control"* was uniformly high among user-participants in all countries, anticipated from the discussion on attitudes towards technology in general.

2 Data Context

The study finds that user-participants provide large swathes of personal information to access the services they want. Many of those interviewed had done so some years previously and do not recall what information about themselves they have supplied.

The study also finds a clear understanding among user-participants that their use of "free" services and applications is bartered against the information they provide to the service provider with whom they interact. If they pay for the software or service, user-participants expect their data to be used only by the service provider. Should there be a dispute they feel a paid-for service offers them greater recourse to a complaints solution than would have been the case if the service they get is free.

"If I paid for the service, then I will fine them. I have a bigger chance of winning [a dispute]. If it is free, then I wouldn't know what to do." (Shanghai)

The study identifies seven variables impacting user-participants' sensitivity on how their data is used (Figure 3):

Figure 3: Variables impacting user sensitivity

"I have a lot of concerns. I have a safety boundary that I cannot talk about [articulate]. A person is not a machine. They have complicated emotions." (Shanghai)

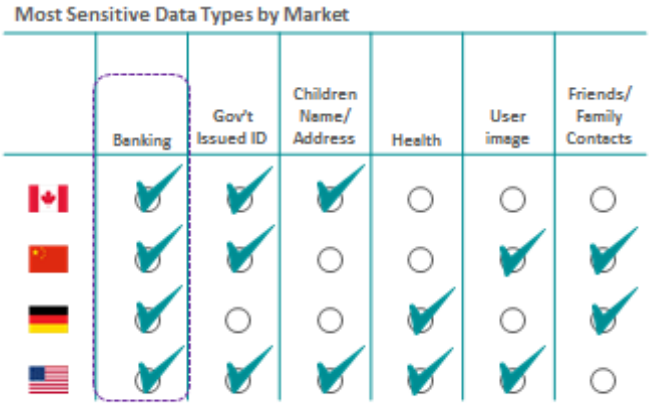
These findings offer preliminary evidence that the current binary approach of identifying a given piece of data as personal and therefore sensitive is not sufficient, and a more nuanced approach that takes into consideration the data context is needed. These variables offer a perspective on the type of regulation or policy framework that might be more suitable, and are discussed in more details below.

2.1 Type of Data

The type of data being accessed or shared is a vital element of the data context – the study finds there are some types of data that are (almost) off limits – "almost" because they are given up only if the circumstance is right and relevant. Financial data are seen as the most "personal" and sensitive by participants. Nonetheless, user-participants understand that the government might need to collect financial data for taxation purposes. Similarly, an employer or online shopping service might be given access to credit card or banking details but other groups or services should not have such access. Importantly, each of the services accessing personal financial information must have a reason to do so. This would not be true of an online dating service, for example, or most social networking sites. So user-participants understand and accept the interconnection between the relevance of access required for certain types of personal data and the legitimacy of the service.

Figure 4: Sensitivity to types of data being shared by country

There are specific types of data that user-participants are sensitive to sharing unless it is relevant to the service provided and yields a benefit to the user



The use of social security numbers or other government-related identification is considered sensitive by user-participants in all countries but Germany. These qualitative data also show that there are other differences among the participants from different countries. For example, health information, considered sensitive in the US and Germany, is not as sensitive in Canada and China, where national healthcare systems are in place, and concerns about existing health conditions impacting insurance costs are not relevant[13].

2.2 Type of Entity

The type of entity users interact with (e.g., large vendors, small start-ups, employers, government agencies), impacts the type of information user-participants feel can be shared. They understand that some entities, such as employers, need to have financial information. This – and other similar comments – shows that user-participants are able to balance their areas of sensitivity with the practical elements of necessity and relevance. In all markets, user-participants feel that the government should not have access to any private data such as personal emails or e-commerce.

"I'm concerned that Revenue Canada will ask [e-commerce site] for my information in order to collect additional taxes." (Toronto)

[13] To investigate the relative importance of these elements and how they relate to other variables, a quantitative study is to be undertaken.

2.3 Trust in the Service Provider

The study started with the hypothesis that trust is a sufficient element in user mental models on personal data. That is, if the user trusts the entities he or she is interacting with, then these entities could have access to any personal data requested, regardless of data context. The research finds that, although trust is an important element within the data context, it is neither a sufficient nor a necessary criterion for user-participants to determine whether they would be willing to share particular types of data. They are not willing to provide more, or more sensitive, information to a service provider simply because they trust it. Other elements of the data context described here are still taken into consideration, including whether the data shared is relevant to the service and/or whether there are perceived benefits in the exchange. However, when trust is lost, user-participants say it can be regained if the service provider is transparent about the issue, fixes the problem, and compensates them accordingly.

In interacting with a service, user-participants distinguish between three different types of trust:

1) ***Service provider*** – how much they trust the entity, and separately, the individuals/employees that work for the entity, to protect the data.[14] Beyond this distinction, users do not differentiate between trust in the "entity", at the broadest level, from trust in a service, an application, a service provider platform, a service provider, or a device.

2) ***Other users of the service*** – how much they trust other users in their network not to misuse the data posted, e.g., their friends or acquaintances may "repost" the data in different and undesirable contexts.

3) ***Content*** – how much they trust the integrity of the content being shared by other users and/or the service providers, e.g., photos, tweets, news articles, etc. This was primarily an issue with user-participants in China, due to their perception of the power of governmental authorities and the frequent account breaches which could lead to the hacking of accounts.

As the original objective of the research was to develop insights into user sensitivity towards personal data in interacting with service providers, and current regulation focuses on this issue, subsequent discussions concentrate on the first type of trust – i.e., trust between users and the entities with which they interact. This trust is impacted by three elements:

[14] In all countries, user-participants brought up their concerns regarding "rogue employees," and clearly differentiate between the overall entity, which may be trusted, and employees of the entity, who cannot be trusted.

1) ***Reputation*** – whether the entity is known through familiarity with the brand, recommendations by word of mouth, lengthy personal relationship, or is a large company.

   "What would make me 100% comfortable? If the risk is perceived to be low. If my friends are using it. And, if it is delivering value to me." (Toronto)

2) ***Location*** – whether the entity has local or national presence, or gives other indications of abiding by local regulations. In all the countries except China, user-participants say this would give them reassurance that there is "some organisation out there" to protect them from harm such as identity theft or fraud. In Canada there was limited reference to the Privacy Commissioner. On the other hand in China there is less awareness that there are national laws in place to protect the use of their personal data.

   "[Companies outside Canada] will have different laws governing what they can do with your personal information. Anything that's outside of Canada, just naturally I'm more sceptical. I don't even really know why." (Toronto)

3) ***Free vs. paid service*** – user-participants understand that if an entity is offering free services, then it must be monetizing their data somehow and that they could be put at risk. If users pay for the software or services, then they expect their data to be used more responsibly.

   "I used [social networking site] for free … If you want the privacy to be completely hidden, it is not possible. If it is free, it means I accept the risk." (Shanghai)

   "[Social networking site] is only free because of advertising. If I don't want advertising then I have to pay for it." (Hamburg)

In general, user-participants have the most trust in entities with which they have financial relationships. These include institutions such as banks, but also include online shopping services.

"It is because they are taking care of my money and you have to have some trust within that organization to take care of your money. That is your livelihood." (Denver)

"Shopping. For me, they are all big companies and they could never afford to get a bad reputation based on data abuse." (Hamburg)

User-participants perceive social media sites to be the least trusted providers. Reasons given include wide access to personal data posted both by the service provider and the community, and difficulty in deleting personal data.

"You can't erase it and they keep track of it. I just changed my login name to my email, and then, I accidentally used my old login name, and it has my profile exactly the way it was. My old login photo. My old conversations to date." (Denver)

2.4 Collection Method

The study used a variety of scenarios to look at how user-participants respond to how the personal data used was collected. A clear distinction was made between data collected *actively* and data collected *passively*, as defined in Section 1.1 above. Overall this research finds that user-participants have a better understanding of active data collection than passive data collection.

With actively collected data, i.e., those data that are directly volunteered by the user-participants, they feel that they have control over the types of data submitted. They also have some limits on the extent to which these data are shared by using different personas, avatars, and online identities, including being anonymous, for certain types of services, such as games. However, when questioned, few user-participants could recall exactly what information they supplied to each service when they first signed up. This did not seem to worry them. Figure 5 below shows the types of data that user-participants submit to different services.

Figure 5: Personal data shared with online services

User-participants say service providers currently have access to both their contributed and observed data, although many cannot immediately recall the amount of data they provided to register for services

User-reported data service providers have access to	Social Media	Shopping	Gaming	Banks	Employer	University
Name	●	●		●	●	●
User ID			●			
Location/ address	●	●		●	●	●
Email	●	●	●	●	●	●
Age/date of birth	●			●	●	●
Demos/life stage	●				●	
Gov't identification				●	●	●
Financial account data		●	●	●	●	●
Relevant user activity	●	●	●	●	●	●
Friends/family	●					
Photos/videos	●					

There are two types of passively collected data:

(1) "Silently" collected data:

Those data that are gathered either:

(a) while the user is actively involved in the transaction, but without the user being aware, e.g., location data while the user is using the mobile phone for a transaction.

(b) automatically without the user being aware of a transaction taking place, e.g., CCTV capture while the user is walking through a train station.

User-participants are mixed in their reactions to the use of "silently" collected data. Their reactions differ based on where it is collected, who benefits from it, and how it is analysed (Figure 6). They are also concerned about whether the collected data is aggregated so that they cannot be individually identified.

Figure 6: Key drivers for participants' reactions towards autonomously collected data

Where it is collected (e.g., jurisdiction)	• **In public**: somewhat expected and out of user's control; user is less likely to be individualized. • **At work**: justified (employer has rights to this data); not likely to be shared outside of the office.	• **At home**: too intimate and personal; reputation and security concerns.
Who it benefits	• **The user**: gives them reward or convenience. • **Society**: data is used to enhance public security (e.g. locating, catching criminals).	• **Service provider**: Users are monitored and data is used inappropriately or sold to a third party.
How it is analyzed	• **In aggregate with other users**: less risk of users being analyzed and targeted individually; safety in numbers.	• **Individually**: concerns that too many details are provided to the service provider (e.g., facial recognition).

"If I am recorded in public places, I'm okay with this data. There is no risk. If there are other people who are hostile, maybe that person has done something."(Shanghai)

"That's kind of a legal issue really. If they put something in your personal home without your knowledge. But the police do that all the time. I don't know. They do it on TV. If it's not community property and it's your own personal space I don't think it's okay. I think you have a right to your privacy." (Denver)

"I'm just a face in a huge crowd. It's not like they've got specific information about me." (Toronto)

(2) Inferred data:

Inferred data are those derived from analytics that aggregate data collected from the user with other available data, including past user transactions or similar transactions carried out by other users. For example, data used for credit ratings. The study finds that user-participants generally are not familiar with the concept of inferred data. Some are uncomfortable with these data being generated; others are concerned about how they are being used, although most agree to usage that would benefit them. Participants are more comfortable if the inferred data are used in aggregate with other users' data.

"The more you reveal about myself (sic), the more possibilities they have to search for stuff about me. I wouldn't want them to know everything about me. " (Hamburg)

"I don't think that is an issue because that is kind of an anonymous – you are not saying this person has a high body temperature and we need to turn up the air conditioning while she is in the room. The right data anonymously would be fine." (Denver)

"It is not your business. I am a very private person and it is not your business if I go get my haircut or I am going for a walk today. I believe that there are still other motives out there. Even though we are using this for marketing purposes only I still sometimes think that there is a hidden agenda or it can be used some other way than what it was initially intended for." (Denver)

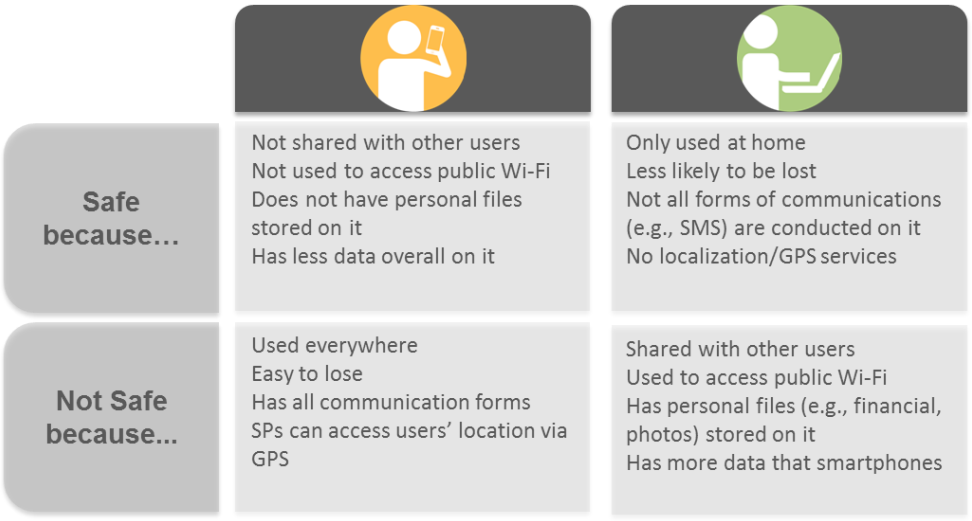
The research finds that many user-participants are concerned about passively collected data. They feel a lack of control and are unclear about how these data might be used and how identifiable they might be as individuals. Many indicated that acceptance of passive data collection would increase under the following conditions:

- they trust the service provider collecting the data,
- they are able to negotiate the value exchanged for their data,
- they are provided with clear insights into how the data is being collected and how it is being used,
- they have control over the types of data being collected, accessed and used.

2.5 Device Context

Much research – particularly among children and young people[15] – has suggested that the mobile telephone is thought to be the most private device, because it is kept on one's person and is not "public". This research does not find such a clear demarcation. Across the markets, user-participants differ in their views about the device from which they think it is safest to access data (Figure 7).

Figure 7: Attitudes towards data access devices



"I feel like if somebody's going to access them, I'd rather have it on my phone (rather than) my computer because I keep a lot more personal information on the computer than I do (on) the phone." (Denver)

"It doesn't make a difference, I do the same activities on both devices (mobile phone and PC). The only difference is the localization services on the mobile phone. They often localize you and I don't have that on the laptop at all." (Hamburg)

2.6 Data Usage

User-participants understand that they cannot get something for nothing. That is, they understand that if they are receiving a free service, for example, their data are probably being used for

[15] http://www.saferinternet.org.uk/Content/Childnet/Safer-Internet-Centre/downloads/Research_Highlights/UKCCIS_Report_2012.pdf

marketing purposes or sold on to other advertisers. But this is an implicit understanding and they see that a benefit is derived on either side.

"The navigation service is free, it provides you a service so I think it's okay that they get your personal information." (Shanghai)

In general, user-participants say they are comfortable sharing personal data when the data are being used as expected or when they receive benefits. They do not want their data used without their knowledge or for the sole benefit of the service provider or an unknown party. Overall, user-participants have negative responses to automated uses of data and they want to be able to decide which type of automated use is permitted. They are concerned that their data might be misused, resulting in potential embarrassment. In Germany and China in particular, user-participants seem concerned about automated data use as they feel their independent ability to make decisions might be compromised. Some of the comments echo sentiments found earlier in the discussion regarding attitudes to technology, when user-participants expressed concerns that automation might lead to a total dependency on technology, obviating the need for their involvement.

"This makes us totally dependent. In 10 years' time you lose your dependency. You don't know how things work anymore. One day I won't know how to do things anymore. I needn't live anymore." (Hamburg)

2.7 Value Exchange

The study finds that user-participants are cognizant of the trade-off between their data and the value ascribed to it (the "value exchange") when they consider whether or not to share their data. Some say they evaluate whether they require the service in the first place and/or if they could get the benefit (such as access to a service) via some other channel without having to share their data (e.g., through an existing relationship). User-participants also recognise that they exchange their data in return for access to a service they might particularly want. We have called this "consent to trust" and it often arises in cases where users consent to trust a service provider and take the risk that their data will be treated in a secure and appropriate manner simply because they want to use the service or are motivated by other benefits. Participants in Canada and China in particular, note that they must trust the service provider in order to use a service.

"I have to accept this passively, because I need it. You need to buy things in a supermarket and you apply for membership cards to get a discount and the supermarket records your consumption." (Shanghai)

A few also say that they feel pressured to adopt the service because they would feel like "a social outcast" otherwise. In the US, participants referred to feeling "resigned" to consenting to trust an entity. Social media are most often talked about in this context: user-participants say they have to belong to a particular social networking site because their peers do; this is balanced against knowledge that their personal data might be "traded" without their consent, either to advertisers or other unknown parties. On the one hand, they want the online recognition the membership of that site confers on them but on the other, they want assurances that this will not lead to the disclosure of data that they consider private (to be shared only among "friends") or even confidential.

"I feel like crap every time. I don't want to do it but I know I have to. You'll become a social outcast. You can't not do it or how are you going to live?" (Toronto)

Nonetheless, most user-participants say they would be prepared to exchange personal data for immediate benefits and for benefits that they can see clearly. These include financial discounts for products or services, better service/improved product, and convenience/time savings.

"I would not give out my weight information normally. But if [store] gave me a 50% discount, then yes." (Toronto)

The research queried whether social good, rather than personal benefit was a motivator for giving out information and found this is not widely considered.

"It depends on what they want it for. If they're environmentally trying to take in information about my energy usage, sure they can track me. If it will help in the future. I think the social benefit is for the good of all." (Denver)

Together, the research found that the seven variables identified above define the data context determining user sensitivity to the sharing and usage of their personal data across all four countries. It will be interesting to assess their relative importance, as well as the differences in these relationships across countries in any follow-up quantitative survey.

3 Understanding Risk

So far we have noted that the user-participant is, in the main, favourable towards digital technologies, has a limited understanding of how personal data are used but has some clear boundaries that make such use more or less acceptable.

Much regulation is based on the concept of detriment or the risk of harm and is designed to offer protection from likely harms. However regulation often seeks to prevent *possible* harm, and so can be excessive and not proportionate to the realization of harms, which are - arguably - measureable[16]. In this context, harm would be the use of personal data in a way that is deleterious to the person whose data are impacted. It is important to note that only one user-participant in this study had had a negative personal experience that had required action.

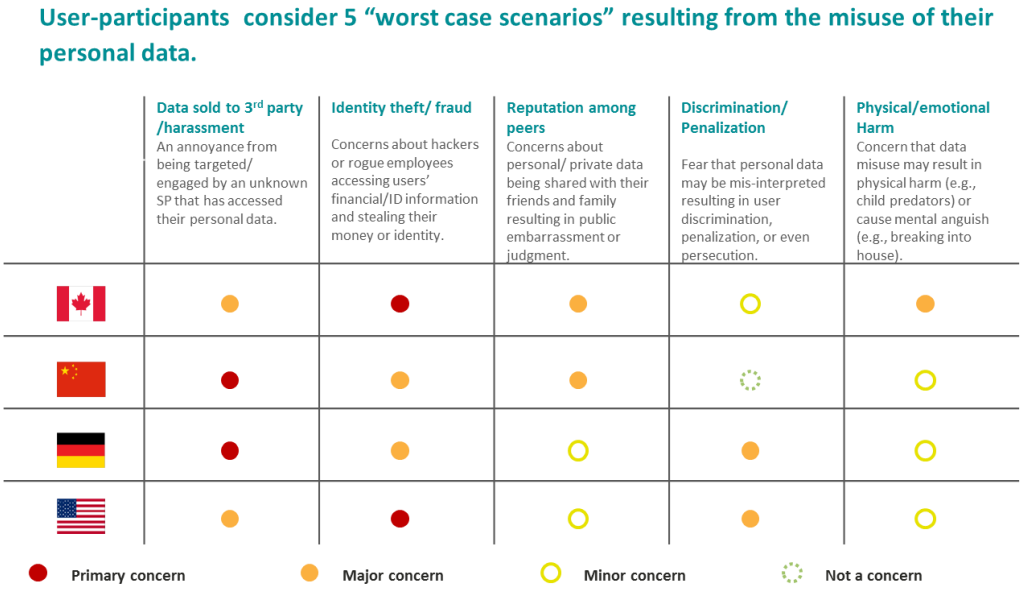
Risk, on the other hand, is a response which takes into account a wide range of relevant factors. Many such factors are culturally-specific; in this study they might include national traditions of regulation. Risk is also determined by personal desires, such as the perceived benefits afforded by the service in question.

The assessment (or perception) of risk by volunteering personal data to a service provider was examined . User-participants say the primary risk is that their data might be sold to a third party. This third party now has personal information - and the consequent offerings made may not be relevant or beneficial. Across all markets, participants describe this as a major annoyance.

"I think what is uncomfortable is that there are companies out there now that know certain things about you that you didn't expect. So, it's a surprise. It kind of seems like manipulation because they know all of these things." (Toronto)

[16] See, for example, Millwood Hargrave, Andrea and Livingstone, Sonia (2006) *Harm and offence in media content: a review of the evidence.*

Figure 8: Perceived risks by country



There are clear differences between user-participants from different countries. Those in North America are most concerned about identity theft and fraud (as well as their data being sold to third parties), while those in China express concern about their reputation being damaged.

"They might send something to your school address. It might affect my reputation. Other people will think I'm weird and people might point at me." (Shanghai)

Other concerns mentioned are issues related to discrimination or prejudicial information being posted online and a sense that they cannot control this:

"I feel like if it's your reputation on (social networking site). People steal passwords and click on these crazy videos. And then they've got a girl with no clothes on their page. Everyone's like, 'Why are you posting all this stuff?' They're like, 'Oh my God my password's been stolen.' You go to get a job later, and they're like 'Oh, a few years ago, you posted a bunch of pictures of all these girls." (Denver)

Few participants express concern about physical or emotional harm that might befall them through the collection of their personal data, although one participant in Denver, when discussing the research scenario involving a woman leaving her home to go to work, said:

"Some criminal mind could be tracking this woman, knowing when her apartment is vacant in case she has millions of dollars' worth of jewellery they can go in and steal. They know where she works. They can cause her all kinds of grief at work. They know her daily routine which if it is not used in a good manner is purely evil." (Denver)

Some user-participants say they avoid the potential for harm by changing their information regularly (sometimes prompted by awareness-raising campaigns or news reports), by using pseudonyms or entering false information or by changing aspects of their online behaviour such as privacy settings.

Few of them read the terms and conditions that they have marked as being "agreed to" when they signed up for a service. Some say the agreement is too long and complicated, or that they want a service sufficiently (perhaps their friends already have it) that they just accept it. They understand this is wrong, and recognise in the main that, by clicking "I accept", they are choosing to take that risk.

"Sometimes it's my own fault. I didn't read the agreement. You make the decision, you have the responsibility." (Shanghai)

Nevertheless there is a sense among many that they are being Canute-like. There has been no harm done. User-participants know they have not been forced into entering their information; they could simply choose not to use the service. But they want to, and more and more services require that they accept an agreement they do not comprehend in order to do so.

In each market, most user-participants say that, while they are concerned about the use to which their personal data might be put, they feel they have no real means of protecting it. They are resigned to the risks associated with sharing such information. To mitigate the perceived risks, many do the following:

- Use several different online identities to manage different personas and data sharing contexts – most user-participants say they designate a specific email address for their important services or high-value transactions (e.g., banking). Some create special identities/pseudonyms for social sites and specific online activities (e.g., gaming, special interest fora) due to lack of trust in these service providers and/or users accessing their data. Yet others create new online identities with faked information when they register for services where they anticipate junk mails.
- Keep accounts separate from family members to keep their information private.
- Volunteer as little data as possible, or input fake information.

4 Perceptions of Accountability

When asked who they think should be accountable for ensuring proper management of their personal data, user-participants think that all stakeholders within the data ecosystem are accountable, including users themselves, service providers, and the government. These perceptions of accountability are summarized in Figure 9.

Figure 9: Perceptions of accountability

"Increasingly regulatory requirements [are] making sure the consumer doesn't get harmed. But, the problem is that [the approach] excludes any responsibility on the consumer; at least in theory. It's still my decision to engage whether something happens or not." (Denver)

4.1 Expectations of Self-Accountability

The study finds that, in general, user-participants have a strong sense of self-accountability and understand that they are accepting a level of plausible risks for the use of their personal data when they accept the terms and conditions of the services they have signed up to, regardless of whether or not they have read them. One participant in the US equated this risk to the risks that one takes when agreeing to obtain a licence to drive a car. If something does go wrong, the user is an accountable party. The concept of "plausible risks" or risks that they have first- or second-hand

experience of is one of the trigger points identified that would motivate users to better control their data.

"Life is full of risks. If you get in the car there is a risk. If you walk there is a risk. And this is our life now. This is the risk we take for all the benefits. I'm not a real big user, but I can see huge benefits to stuff." (Denver)

"I didn't read the agreement. You make the decision, you have the responsibility." (Shanghai)

A few participants in China and Germany say that the agreements outlining the terms and conditions for a service only serve the interest of the service provider, and do not address their own rights. They would like instead to have a two-way contract in place that outlines both the service provider's obligations and the user's rights.

When questioned, user-participants indicate some willingness to receive information and become more educated, both on how service providers use their data and by reviewing the data already held by service providers. Nevertheless, they also question how much effort they would commit to these activities in reality. Some user-participants feel government should be regulating appropriate uses of data, and that it is not their personal responsibility. These are valuable inputs into the transparency requirements that are currently part of privacy regulation discussions.

"I would review the information companies have on me initially, and then only again if there is any issues." (Toronto)

"They are accountable for keeping it safe, but you are responsible for giving it to them." (Denver)

4.2 Accountability of Service Providers

As the last quote indicates, although user-participants say they are willing to be accountable for the type of data they put online, they expect service providers to treat their data responsibly, keeping their data safe and using it in a way consistent with their expectations. Existing local regulations are expected to be complied with and service providers must be accountable, taking responsibility and action should they not be compliant or commit a breach of trust.

For vendors, fulfilment of these expectations, with regulatory enforcement, would lead to greater trust. This finding reinforces the earlier discussion on trust, where location of an entity is important

because of the perceived existence of a formal dispute resolution procedure through the enforcement of local regulations.

It is noteworthy that some user-participants in the US feel that service providers should not be held accountable for data breaches if appropriate precautions and measures have been taken to keep users' data safe. The same sentiments are heard in China where participants are aware of data security breaches.

"You trust the bank. They had credit card information stolen; it's a huge entity. You'd think this would never happen to [bank], but it happened, so you can't blame them because they went through all their security precautions, and someone just smart enough just kept working at the code and working at the code. It's just like the flu viruses." (Denver)

Expectations of employers regarding accountability for personal data are more limited. Employers act as service providers in that they facilitate access and so might be thought of as "intermediate service providers". User-participants talk of the accountability of employers mostly in relation to the appropriate stewardship of employee data. Discussions of user sensitivity to sharing personal data with their employers reflect interesting cultural differences regarding whether employers should have access to non-work data. In Canada, user-participants indicate that users are accountable for using employer-owned resources for personal activities (e.g., using their work PC for personal email), and that employers would have rights to these data. User-participants in the US indicate that employers only have rights to work-related data and not to any personal data, regardless of whether work equipment or resources were involved.

All participants feel that their personal data should be removed if the employee leaves the employer.

"If I am still in the company it's okay but if I quit my job, the company should delete my information." (Shanghai)

4.3 Accountability of Government and Intermediaries

Integral to perceptions of accountability is the notion that entities are monitored and/or regulated by the government. The notion of dispute resolution and the possibility of redress came up as elements impacting the degree to which users trust service providers. User-participants express clear expectations of the government to establish, administer, and enforce data protection regulations, but not to monitor, access, or control personal data.

In Canada the Privacy Commissioner was mentioned as someone who would have powers to act should there be a misuse of personal data. Other participants also express a sense of security because of their conviction that there are national structures governing the use of their personal data.

"The government can make sure that my information is not being shared." (Toronto)

"I don't feel like investing my time to look at these things. I fully trust that someone has verified the legal character." (Hamburg)

The tension expressed between the need to rely on the government for enforcement and the fear that such power may be abused leads many user-participants to suggest the creation of non-governmental intermediaries with enforcement authority who would ensure appropriate data usage and would safeguard their data.

"I want a third party to manage the data. There must be a high requirement for a safe data centre. Your information is saved at the centre, not the store. It has a powerful defence against hackers and won't disclose information to third parties." (Shanghai)

These intermediaries or third parties could process complaints and mediate between users and service providers. They could set quality of service standards and act as enforcement authorities ensuring compliance with whatever regulations are in place. The intermediaries would also manage user data across multiple service providers, ensuring a level of consistency in expressing personal preferences for data and providing assistance to users in the management of their personal information.

There are two things to note –Canadian participants do not mention the concept of an intermediary (while there is some mention of the Privacy Commissioner, already noted), and user-participants do not want to pay for such an intermediary. It is felt that the funding of such a body should be by the service provider who benefits from the data collection or the government (which benefits less directly, perhaps through economic good).

5 Expected Data Principles

When asked to identify the data management practices that they would want in place to address their concerns when providing personal information, user-participants across all markets identified five key principles, noting that accuracy in the transmission or transference of their data is a key underlying assumption:

"Accuracy is important if the information is being sent out." (Toronto)

1) **Control** – access to a graduated level of control over how their data will be used throughout their relationship with a service provider (e.g., initial opt-in/out, ongoing permissions for data access, automation approval, enable data sharing, delete data upon termination of relationship, request data for portability), even if some user-participants question the amount of time/effort they would actually spend to exercise this control.

   "But it comes down to time, how much time do I have? Do I really have the time to sit there and go back through all that data and decide – 'I don't want them to have it, but that's okay for these guys to have it.'?" (Denver)

2) **Transparency** – greater transparency and clearer insights of how their data are being used and by whom; user-participants accept they have no clear idea how such information should be delivered to them or whether they would use the information.

   "It's important that it's very open. Make plans transparent. [Show why the service provider] wants the data, what it will do, the purpose of collection, the objective and the benefit of the person who is receiving the data." (Hamburg)

3) **Value exchange** – user-participants expect to receive a benefit (e.g., financial reward, enhanced service, convenience) which they perceive should be of comparable value to their data. Although some participants mention social good, personal benefits are the most appealing.

   "Tell me if you want to sell [my data]. Then I should get a cut of it." (Toronto)

4) **Accountability** – expectations of the accountability of all stakeholders in the data ecosystem as discussed in detail above.

5) **Protection/Enforcement/Redress** – user-participants want assurance of regulations that can deter abuse of their personal data, that these regulations are enforced, and that the due process for dispute resolution is clearly defined.

6 Key Differences by Market

As already noted, this qualitative study was conducted in four countries (Canada, China, Germany and the USA) chosen to reflect the differing approaches to privacy regulations and differing social and cultural norms. Part of the interest in this research was to identify commonalities among peoples in different regions – how their mental models compare and their expectations of the way in which their personal data might be used. As the data above show, there are general similarities but there are also differences, drawn together below.

Figure 1: User type characterisation by country

Four personal attributes uncovered in this research make up the dimensions of user types, including:

Personal data awareness	Low	Low	Mid	Mid
Trust in government	High	Low	High	Low
Perceptions of own accountability	High	Low	High	High
Desire for control	High	High	High	High

Figure 1 above (repeated here) shows that user-participants in each of the territories generally have limited awareness of how their personal data might be used and of the data ecosystem within which they operate[17]. Differences are noted when measures of accountability are examined. In Canada and in Germany there is trust that the government will have set rules in place that will provide a framework governing the use of personal data:

"It would be nice if the government verifies [the usage of my data]. Something like "state examined" so you see that someone's really looked at it. Like an independent foundation testing consumer products." (Hamburg)

In China, by contrast, user-participants talk of data breaches that have occurred and seem to place less trust in the government whereas in the US, user-participants are concerned about the government extending levels of control over their data so that users have decreasing authority.

"I don't trust them to regulate. I think because they always take it a step further. They may come in and say 'Well, we're going to pass this law because it'll help something', but then they'll realize they can push it further." (Denver)

The level of trust in government is also related to attitudes towards the location of the service or service provider. If the user-participant trusts their government to provide a framework for the management of their personal data through laws etc., as was expressed more often in Canada and Germany, then user-participants are more likely to express a desire for the service to be located in that country and/or to abide by national laws.

There are a few differences noted between the four territories in the type of data that is considered most sensitive. While financial/banking information is thought by all four markets to be the most sensitive type of data, identity cards or other government-administered identification systems are considered sensitive in all the countries studied, bar Germany. The need to protect children is mentioned more frequently in North America, while access to one's contacts is mentioned in both Germany and China. In China too, there is concern about the use of one's personal image in inappropriate circumstances.

These findings show some correlation with the risks that are perceived to be associated with putting personal data online. So in North America where financial and identity-related materials were thought to be very sensitive there was most concern about fraud and identity theft. In China and Germany, user-participants talk more generally about third parties having access to their data and sending them information or linking them to sites in which they have no interest or are not relevant to their needs.

"If it is accompanied with a good intention and keeps the information to itself and it wants to make the customer happy, then it's okay. But, if it sends the data to someone else, it's not funny anymore." (Hamburg)

A closer examination of attitudes towards accountability underlines the difference between China and the other countries (within this qualitative research). In China trust in the government and personal accountability with reference to personal data management are lower than in the other

territories and greater obligation is placed on service providers, particularly large global providers, to manage data securely. Participants in China are also most likely to call for an intermediary to act as a "storage space" and quality control medium between the user and the service (this intermediary may be backed by government so that legal redress can be sought).

"I want a third party to manage the data. There must be a high requirement for a safe data centre. Your information is saved at the centre, not the store. It has a powerful defence against hackers and won't disclose information to third parties." (Shanghai)

Figure 10: Perceptions of accountability by country

7 Conclusion

7.1 User Mental Models

By exploring the mental models that user-participants have about personal data and their uses, this study clearly indicates that participants reflect many of the issues that regulators are grappling with in their consideration of the safeguards that should be applied to the use of personal data. However, the mental models also reflect a more nuanced approach towards personal data, based on the recognition of an integrated data context and assumed data management principles, than the current binary approach taken by regulators.

User-participants express a desire to exercise control over what happens to their personal data and how it is used, in particular by third parties. It is the onward use of data – information passed on to third-party services – that creates most discomfort, and user-participants feel they have no choice but to trust the first (or initial) service provider not to misuse their data. The degree of trust is based on many variables and, in the countries studied, a degree of regulation is expected as a backstop power. Some participants consider that an intermediary would be an effective mediator between the various stakeholders; in this case, the user, the service provider and the government or statutory regulator.

User-participants also accept that they often do not seek to exercise the control they can have once they enter into an online transaction with a service provider, and so cede much of that control in return for the service being offered. However the study also demonstrates that user-participants attach graduated levels of sensitivity to different data, and differentiate between types of online services (especially those that are free at the point of use and those that must be paid for). These, and other, variables lead them to expect different uses to be made of the data and can affect their decision about the amount and type of personal data they will provide. Thus the binary definition used in much regulation of "is/is not personal data" is not sufficiently discriminating to be attached to emergent services that could be restricted by a simplistic definition, and a graduated or progressive system of control should be provided to the user.

While many participants indicate they do not read online agreements such as terms and conditions and admit that they are not willing to make the effort or time to do so, there is a desire to have such information available in a simple and accessible form. They want to have increased self-determination in how their data are used. This requires access to relevant tools and structures that

enable users to specify levels of permission for the use of their personal data within the data ecosystem. Such tools and structures include

- simple and easily accessible guidance which explains the choices the user is making when personal data is volunteered;
- clarity about the ways in which data will be used (for both free and paid for services);
- transparency around issues of portability and deletion of personal data;
- adequate systems of accountability in place to ensure compliance so that breaches are dealt with fairly and swiftly; and
- (for several user-participants) intermediaries that keep track of and manage user preferences consistently across multiple service providers.

The findings show that a perception that users do not exercise control or cannot manage their personal data, and so are at significant risk of harm through the misuse of their data, is incorrect. While user-participants themselves agree that there may be a risk and are aware of their potential vulnerability, they regulate what information they give to different types of services and what they do online. This is demonstrated in a number of different ways:

- they exercise control over the data they input and make judgements about what they disclose based on the benefits they derive from that service;
- they manage several online personae with different functions or purposes (e.g., spam, financial, work, friends);
- they keep accounts separate from other family members and work;
- they volunteer the minimum data required to access a service, or enter false information;
- They use pseudonyms to maintain anonymity.

"If you use one pseudonym everywhere, it's easier to abuse it if someone finds out the password." (Hamburg)

Many also take other measures. For example,

- following public recommendations, they choose to refrain from sharing sensitive data (e.g., government identification data, financial information);
- they research the reputation of new service providers;
- they set and re-set privacy settings on social media sites.

"Reading something about identity theft; you shouldn't do this or you should do that. I will go in and check and see if I am matching what they say you should do." (Denver)

These findings point towards a system of controls which is triggered as the nature and quality of data disclosed by the user alters. While it is unclear to what extent they would use such a system, user-participants ask for:

- **Trust and accountability** – user-participants demand accountability from all entities involved in the personal data ecosystem (government, service providers and users etc.), and recognise their own accountability in this context.

  "Increasing regulatory requirements [are] making sure the consumer doesn't get harmed. But, the problem is that that excludes any responsibility on the consumer; at least in theory. It's still my decision to engage whether something happens or not." (Denver)

- **Graduated control** - at each level of disclosure a different level of reassurance or increased benefit should be given – e.g., initial opt-in/out decision, granting permissions for data access, automation approval, enabling data sharing and requesting that data be deleted when the relationship is terminated.
- **Value exchange** – in exchange for their data, user-participants expect to receive a benefit (e.g., financial reward, enhanced service, convenience) which should be of comparable value to their data.
- **Transparency** – user-participants call for clear and precise, not more, information (although they are unsure how these should be delivered) about the entities which have their data and how it is being used, so that they can exercise control should they wish to.
- **Protection/enforcement/redress** – user-participants want the assurance that any regulations mandated by the government will be enforced in order to deter the abuse of their personal data by service providers, and that there are clear dispute resolution procedures in place.

7.2 What Does This Mean for Regulation?

Policymakers and regulators recognise the importance of the digital world in providing economic and social benefits. The goal of regulation around personal data is to ensure that no harm is caused to the user. However, regulation should not seek to guard against every *possibility* of harm, rather than the *likelihood* of harm. The danger is that such a blanket approach will stifle or at the very least 'chill' innovation, to the detriment of business investment, economic growth and social benefits

While personal data management – the subject of this study– is an issue of concern for regulators, it is difficult to locate under any one single heading within the regulatory framework. Much current

user-facing regulation is not about the passive collection and aggregation of personal data but about active data collection. This is, as we have seen, only one way in which personal data are used – "big data" specifically (passive data) is where the rapid growth is. From a user perspective, the study finds little understanding of passive data collection and, when it was emphasised to them, some uncertainty about how it might be used or is useful. It may fall therefore, to stakeholders such as industry and regulators to explain how big data are collected and to help create policy frameworks that are cognisant of these developments and can respond effectively and flexibly while not impeding the benefits that may result from their implementation.

This report shows that there are multiple regulatory issues that must be addressed. Current regulatory frameworks do not consider all aspects of personal data collection and management. Regulation has addressed them primarily through privacy regulation, concentrating on protection from harm, rather than a calibrated set of responses that reflect the different types of data collected and the uses to which they are put – and the user mental models that understand these differences. The research found:

- there are variables that affect the data context and these in turn, impact user sensitivity regarding personal data.
- participants had an understanding of the potential for harm and took measures to mitigate against risk.
- participants had strong requirements of accountability and the levels of responsibility of different stakeholders, including themselves.

The data context and aspects of self-regulation noted (such as using multiple passwords depending on the service being accessed) are key to users' mental models, and define the parameters needed for user-participants to experience a "safe environment" for their data. The elements of accountability and responsibility reflect an expectation that there will be principles or data management practices in place and that data will be secure. This aspect offers users a base level of trust, which in turn impacts the data context. Co–regulation (the inclusion of intermediaries into the personal data management environment) is not dismissed, and preferred in some cases, as it removes a sense that too much power could rest in governmental organisations.

The study shows that the issues around personal data management are not linear and regulation cannot assume that all data are the same and carry the same weight between individuals. They change, based on factors such as culture, social norms, gender and awareness.

A further complication in the regulatory debate is that "privacy" can become conflated with "security" (in particular national security) which is at the other end of the social and political spectrum in a

debate about personal data management. However the technology used is the same and the difference is in rights of access. This highlights again the fragmentation of regulation in this area, and the need for regulatory principles that are high level and flexible enough to address the different aspects of data management policy.

Another conundrum is the disparity of regulation – and law – from country to country (even within a single economic zone such as the European Union). Data flow across borders in a digital environment and regulation will need to provide for consistency and some basic principles that can be adhered to – especially by multi-national organisations/service providers. The fact that data flows across national boundaries and is increasingly global in terms of storage is not raised by participants in this study but it is a prime consideration for regulators with their own national structures and for service providers who offer transnational services and storage in the cloud, for example.

Principles, not rules - In those countries with strong sectoral regulatory backgrounds (such as broadcasting content standards in the UK), rules have given way to principles, with law as the backstop power ensuring a base level of protection. This allows for differentiation between services as well as flexibility in enabling new business models and technology evolution, while building a common denominator level of oversight. In some fields a self-regulatory approach is taken – the major mobile telephone operators have developed a privacy code which sets out some "high level privacy principles," although the code does not explain whether a common compliance procedure and monitoring system is in place. This is a model that might be considered, although it may be a step too far in a developing and still largely untested market.

Co-regulation - The study suggests that – for some participants - co-regulation might be acceptable where an intermediary organisation offers a role as a specialist in the field to ensure compliance but has legal backstop powers to support it. This suggests voluntary codes of conduct backed by powers of enforcement, such as in broadcast advertising in the UK[18].

Education - Education/ awareness/communications literacy are all aspects of a vital knowledge base that can be built to ensure that users understand better the personal data ecosystem and, more generally, how they might interrogate it. Creating a knowledgeable user base would be preferable to having restrictive processes such as opt-in models imposed on all users. Many mobile

[18] http://www.asa.org.uk/Regulation-Explained/Control-of-ads/Co-regulation-broadcast.aspx

operators now offer leaflets at point of sale or oblige parents to make active choices to enable access to the open internet when buying mobile devices for their children[19].

Global coordination - The global nature of data flows was not raised by participants in the study. Regulators however, have to consider this in the context of personal data management. There are bilateral agreements in place (such as the Safe Harbor agreement) but nothing that is global or truly transnational.

Shared permission - The study suggests that user-participants recognise that personal data management is not solely an issue for the protection of the individual to whom those data relate. It is a shared, not exclusive, relationship. The individual enters into a contract with the service provider and has to take responsibility for the data volunteered. The obligation of the service provider is to ensure that the user understands this and is given appropriate mechanisms to indicate this is so.

Flexibility - Finally there are characteristics of personal data that will always challenge regulators:

- There are no clear norms and the context in which personal data is collected is shown to be important and variable. Any concept of personal data management that assumes protection is required for all data is too rudimentary;
- The exponential growth of data;
- The ease of distribution, across boundaries and borders;
- The ease of duplication and re-use;
- The ability to fuse data sets together.

These characteristics argue for a system which is transparent to the user and indicates in a clear and accessible form how his or her data are to be used. It needs to be based on a set of data management principles, such as accountability, and appropriate and effective enforcement procedures for breaches that are not illegal but may be harmful.

While these data are based on findings derived from research among users of technology, they are not all "early adopters" or high-level users which suggests that many users of technology do take action to protect themselves from harm, if not from all risk of harm. This study suggests that users are willing to participate in a cooperative process but want to do so in a way that is instinctive (in the way that their use of services is) and easily accessible.

[19] https://wbillpay.verizonwireless.com/vzw/nos/uc/uc_content_filter.jsp

Further research will be done to test these models, and to examine the differences that were noted, resulting from differing regulatory structures, social and cultural norms. This examination of difference, as well as commonalities, is important because data flows across global borders and regulatory principles and practices therefore need to converge and adapt to reflect this, protecting users from harm wherever they are.

APPENDIX A – Methodology

SAMPLE

Participants in this study came from Canada, China, Germany, and the US, chosen on the basis of their different data protection regulatory regimes and cultures. They ranged in age from 21 to 60, with an even mix of male and female, single and married, and include full time students and a variety of professions.

In total 76 participants took part in the research, recruited as online service users and against a battery of questions which tested their personalities on the following four profiles:

- technology enthusiasts who must always have the latest technology,
- technology users who are constantly on the move,
- social networkers who need to stay connected with friends, and
- technology pragmatists who use technology to manage their life and home activities.

The research was carried out between May – August 2012.

METHODOLOGY

The research used a mixture of focus groups (of between 6 and 8 participants) and individual depth interviews. All the research was undertaken by IPSOS UU.

At the start of each session, participants were asked about the online services they regularly accessed, the identities used for each service, the personal information provided, and the devices used for access. Scenarios were then provided to be used as context for participants to answer questions about their mental models with regard to different uses to which their personal data might be put. The scenarios used included the following:

- The concept of value exchange: examining what personal data would be exchanged in return for benefits and how far beyond the individual these data could go.
- The use of automated data: data collected on personal habits to offer benefits that might be of interest to an individual.
- The use of passive data collection: to create societal benefits that would positively affect the environment at large.
- The use of personal data: within the work environment to improve one's working life.

APPENDIX B – Detailed Research Survey Results

See:

http://iicom.org/resources/open-access-resources/cat_view/8-iic-resources/3-open-access-resources/29-personal-data-management