Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

02-Apr-2013

English - Or. English

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY
COMMITTEE FOR INFORMATION, COMPUTER AND COMMUNICATIONS POLICY

Cancels & replaces the same document of 29 March 2013

Working Party on the Information Economy
Working Party on Information Security and Privacy

EXPLORING THE ECONOMICS OF PERSONAL DATA: A SURVEY OF METHODOLOGIES FOR MEASURING MONETARY VALUE

JT03337330

FOREWORD

This report takes an initial look at methodologies to measure and to estimate the monetary value of personal data. Personal data is creating economic and social value at an increasing pace. Measuring and estimating the value being generated, however, is difficult. This is because not only is a huge amount of data being generated, but personal data is used in many different situations for numerous purposes. Studying the value of personal data begins with comparing methodologies for assigning monetary values attached to personal data.

In preparation for work in this area the Working Party on the Information Economy (WPIE) and the Working Party on the Information Security and Privacy (WPISP) jointly organised a Roundtable on the Economics of Personal Data and Privacy, held on 1 December 2010. Three background papers were commissioned for the roundtable and a report of the proceedings from the roundtable is available.[1]

This paper was developed by the OECD Secretariat (Christian Reimsbach-Kounatze, Taylor Reynolds, and Piotr Stryszowski) for consideration by the WPIE and WPISP. It was declassified by the Committee for Information, Computer and Communications Policy (ICCP) by a written procedure concluding in February 2013.

It is published under the responsibility of the Secretary-General of the OECD.

[1] www.oecd.org/sti/privacyanniversary

TABLE OF CONTENTS

FOREWORD 2
TABLE OF CONTENTS 3
MAIN FINDINGS 4
Introduction 7
Technological and regulatory context 7
What are personal data? 7
Purpose and scope 9
The personal data economy: Mapping value chains and business models 10
Methodologies for estimating the value of personal data 18
Financial results per data record 20
Market prices for data 25
Illegal markets 27
Surveys and economic experiments 29
Revealed willingness to pay to protect (insurance) 32
Conclusions and next steps 33
A need for better data and collaboration 33
Understanding the regional context 33
The utility of case studies to understand effects 33
Potential non-linear returns means network effects 34
Capturing producer and consumer surplus will be difficult 34
Markets where individuals control and sell their own data are evolving and will provide insights 34
REFERENCES 37

Boxes

Box 1. Personal data analytics, distribution and usage: The case of online advertising 14
Box 2. Platforms for smart phones 15
Box 3. Data brokers 16
Box 4. Use of personal health records 18
Box 5. Data breaches and their economic impacts to firms and consumers 29

MAIN FINDINGS

Business models that rely on personal data as a key input are increasingly common. Firms are now able to attain significant market valuations by employing business models predicated on the successful use of personal data within the existing legal and regulatory frameworks. The rapid emergence of personal data as an asset in business processes, enabled by the development of ICT, calls for factual, empirical analysis to understand its economic mechanisms. This analysis may enhance policy makers' understanding of issues related to the Internet economy, in particular for policy makers responsible for rules governing the collection and use of this data. It also lays an initial foundation for future research on the societal and economic impact of personal data usage.

The collection and use of personal data is governed by regulatory frameworks in operation across the OECD. These frameworks are evolving to address the current context, with data privacy reviews underway around the world, including at the OECD. The Terms of Reference for the Review of the 1980 OECD Privacy Guidelines highlight the change in scale in the volume of personal data being collected, used and stored, as well as the value of the societal and economic benefits enabled by new technologies and responsible uses of personal data. Establishing a better understanding of the role of personal data in the economy and society can provide context to policy makers working on data privacy reviews.

This report provides a preliminary survey of methodologies for measuring and estimating the value of personal data from a purely monetary perspective (i.e., without taking into account the indirect impacts of the use of personal data on the economy or society — a caveat that will be discussed below). It looks at a range of measurement and estimation techniques and identifies the main benefits and drawbacks of each approach. The current report should be seen as a useful step towards a broader analysis that also addresses the issues of value from the perspective of the society and individuals.

The current analysis seeks to contribute to a broader understanding of the developments in the Internet economy that are driving innovation and growth. It is clear that the Internet is not simply an efficient tool for acquiring and assembling personal (and other) data, but also a key platform for many emerging, innovative applications that transform the data into valuable services for consumers and businesses.

All valuations coming out of the methodologies should be used cautiously, with the understanding that the monetary estimates of values will be context dependent, as is discussed below. In addition, relying solely on one specific approach will likely lead to biased results. Likewise, there are difficulties in determining precisely what are personal data, and in comparing different proxies for personal data such as "records" or "users." However, understanding the values emerging from different approaches can help provide insight into the broad range of monetary valuations attributed to personal data by markets.

One approach to determine the monetary value of personal data is to examine **market capitalisations, revenues or net income per individual record** for firms whose business models are based primarily on personal data. Market capitalisation data leads to valuations that can fluctuate considerably. For example, the implied market capitalisation per Facebook user has fluctuated between USD 40 and USD 300 at different times between 2006 and 2012, with a recent valuation in May 2012 approaching USD 112 per

user. These fluctuations appear to be influenced to a large extent by other economic factors over the time period, and not solely by the monetary value of the underlying data.

Revenues or net income per record/user seem to be a more stable measure of the annual market value of personal data. For example, Facebook and Experian, two companies whose business models are based on personal data, have annual revenues per record/user of roughly USD 4-7 per year. While imprecise, the data can still provide a useful point of reference, although it is worth noting that only at the level of net profit per record are we actually measuring added economic value.

The most direct way to approach the value of personal data is to evaluate the **market prices** at which personal data are legitimately offered and sold. The values are imprecise because they only represent the price of data sold in a specific context to one participant, and do not reflect the total "earnings" of the data over time. They do, however, provide a market-based measurement based on the intersection of supply and demand. At the time of writing, examples of prices in the United States for personal data ranged from USD 0.50 for a street address, USD 2 for a date of birth, USD 8 for a social security number, USD 3 for a driver's license number and USD 35 for a military record. These are only estimates but provide some insight into the relative market values of different pieces of personal data.

Another methodology to look at the monetary value of personal data is via an assessment of the economic **costs of a data breach**. This includes measures of the cost to individuals who have their identities stolen or the costs to firms when there is a data breach. The costs associated with the loss and misuse of personal data may also give some indication of its value, although reported figures vary widely. These approaches do not measure the value of the underlying data but rather the monetary cost of a breach on a per-record basis. The TJX data breach, for instance, forced TJX to set aside USD 118 million to cover costs and potential liabilities in fiscal year 2008 (USD 1.18 per record). It does not, however, cover loss in reputation, impact on the brand, and other indirect and opportunity costs. A more recent example is the security breach of the Sony's PlayStation Network and Sony Online Entertainment in 2011 which resulted in the exposure of 103 million records. According to Sony executives, this data breach will cost the company at least USD 171 million (USD 1.7 per record).

Yet another way to put an economic value on personal data is to run **economic experiments and surveys** that extract the price firms would need to pay individuals to give up some of their personal information. In recent years, several experimental studies have attempted to quantify individual valuations of personal data in diverse contexts. Even though research in this area is still in a preliminary stage, two general messages can already be extracted. First, people tend to differ with respect to their individual valuation of personal data (*i.e.,* the amount of money sufficient for them to give away personal data) and their individual valuation of privacy (*i.e.,* the amount of money they are ready to spend to protect their personal data from disclosure). Second, empirical studies point out that both the *valuation of privacy* and the *valuation of personal data* are extremely sensitive to contextual effects.

Another way to put an economic value on someone's personal data is to understand how much the individual would be willing to **pay to protect** that data in the form of **insurance**. This can also be seen in markets in the form of insurance policies to protect against identity theft. Experian, a data broker, sells an identity-theft protection service called ProtectMyID for USD 155 per annum in the United States. It is interesting to compare this figure with the average revenue per record (USD 6.42) and market capitalisation per record (USD 19.24) for some perspective on the difference between measures. It should be highlighted that there is some debate about whether services that offer to protect and rectify the theft of personal data are actually effective, and therefore can be viewed as a valid instrument of measurement.

Better data is needed to understand the economic value of personal data. It is in the joint interest of policy makers and firms throughout the personal data value chain to work together to ensure that personal

data policies are efficient. Firms have a good understanding of the number of records they hold and should have a strategic view of their monetary value. Policy makers and firms should work together more closely to better understand the potential value of personal data and support efficient policy making.

Estimations of the monetary value of personal data are highly context dependent. This makes any focus on the macroeconomic effects much more difficult because of a lack of harmonised data and measurable impacts over time. Focusing on case studies geared at understanding the effects in a range of specific contexts such as health care or transportation may yield more interesting results. In addition, in most cases the presented methodologies are illustrated with examples that are based on US data. Therefore, caution is needed in extending these illustrations to other countries, whose business environment and regulatory frameworks may significantly impact the results.

Even if the monetary values of personal data can be extracted, they would not cover the full economic and social benefit — both personal and for society at large — that can be derived from the data. An analysis that omits the consideration of consumer surplus will likely underestimate the true social and economic benefits of personal data, since it is not captured in market prices.

There are new developments on the horizon that could help improve our understanding of the monetary value of personal data. New "data lockers" allow users to contribute and control data sharing with third parties in exchange for a portion of the proceeds from the use of their data. These data exchanges could provide new market-based estimates of monetary values, and potentially improve transparency about how data is collected, sold and used.

Introduction

The outcomes from the 2008 Seoul Ministerial on the Future of the Internet Economy continue to provide direction to the work of the Committee for Information, computer and Communications Policy (ICCP). The Ministerial Declaration calls on OECD to analyse developments in the Internet economy that are drivers of innovation and growth. It is clear that the Internet is not simply an excellent tool for acquiring and assembling personal data but also a key platform for many emerging, innovative applications that can transform the data into valuable services for consumers, businesses, and researchers. Governments are also looking for ways to support innovation in the economy and many of the most popular innovations that have emerged over the previous several years (such as social networking) have been predicated on personal data.

The Ministerial Declaration also calls for an assessment of OECD instruments, including the 1980 Guidelines on the Protection of Privacy and Transborder Flows of Personal Data ("Privacy Guidelines"), in light of changing technologies, markets and user behaviour and the growing importance of digital identities – each of which affects personal data. The terms of reference for the review of the Privacy Guidelines highlight the change in scale in the volume of personal data being collected, used and stored, as well as the value of the societal and economic benefits enabled by new technologies and responsible uses of personal data.[1]

Technological and regulatory context

In the late 1960s and 1970s, significant developments in computing and communications technologies created a sense of urgency about threats to privacy, out of which emerged a set of policy principles that continue to govern uses of personal data today. The outcome of that period, which produced the OECD Privacy Guidelines, is now reflected in privacy legislation widely enacted across the OECD and beyond.

The last decade has witnessed extraordinary developments related to data-storage capacities, high-speed networks to move data and a rapid growth of computational power. Significant changes in the volume and uses of personal data have been enabled by improvements in the ability to collect, store, aggregate, link, analyse and transmit personal data. The shift from analogue to digital technology has led to a much greater capacity to store and share pictures and video. Mobile devices enable the routine collection of geolocation information that locates individuals in time and space. Sensors used in health, environment and energy sectors produce data that can be linkable back to individuals. And much of this data is made available globally, supported by communications networks that permit continuous, multipoint data flows.

Policy makers today are now engaged in processes to review and update privacy laws and frameworks across the OECD. As was the case in the 1960s and 70s, it is technological developments – and the changes to personal data processing they enable – that are driving the renewed attention to the rules and institutions set up to govern the collection and use of personal data and protect individuals' privacy.

What are personal data?

The key term for the analysis in this report is "personal data". This report uses the definition contained within the OECD Privacy Guidelines: "any information relating to an identified or identifiable individual (data subject)." This is a broad concept, which includes, by way of example, the following types of personal data:

- User generated content, including blogs and commentary, photos and videos, etc.
- Activity or behavioural data, including what people search for and look at on the Internet, what people buy online, how much and how they pay, etc.
- Social data, including contacts and friends on social networking sites;
- Locational data, including residential addresses, GPS and geo-location (e.g. from cellular mobile phones), IP address, etc.
- Demographic data, including age, gender, race, income, sexual preferences, political affiliation, etc.
- Identifying data of an official nature, including name, financial information and account numbers, health information, national health or social security numbers, police records, etc.

Research literature has further categorised personal data in different ways. For example, Schneier (2010) has developed a taxonomy of personal data, using social networking sites as an example, and differentiated six types: service data, which is provided to open an account (e.g., name, address, credit card information, etc.); disclosed data, which is entered voluntarily by the user; entrusted data, taking as an example the comments made on other people's entries; incidental data, which is about a specific user, but uploaded by someone else; behavioural data, which contains information about the actions users are undertaking when using the site and may be used for targeted advertising; and inferred data, which is information deduced from someone's disclosed data, profile or activities (OECD, 2010).

Personal data is also often categorised according to its use. It is common to distinguish between data collected by a particular entity for use during the current Internet session, and that which is stored for use and analysis over time and/or sold on to third parties (FTC, 2009) – often referred to a first-party and third-party use, respectively.

Another common distinction is that between personally identifiable information (PII) and non-PII. PII is data that directly identifies a person, whereas non-PII is often considered not to be identifying of itself, although the distinction is not always clear. For example, PII typically includes name and address information, social security, health or other unique identifying numbers, health records and financial information. Non-PII typically includes data relating to what search terms we have used, which websites we have visited, what we have purchased online and how we paid for it, etc. There is less certainty about the categorisation of some other types of information (e.g., GPS position, IP address, etc.), and there is also less certainty as analytical methods improve and enable greater (re)combination of pieces of information that are of themselves non-PII, but which, in combination, may allow individuals to be identified.

Distinguishing between personal and non-personal data is becoming increasingly difficult. "Once any piece of data has been linked to a person's real identity, any association between this data and a virtual identity breaks the anonymity of the latter." (Narayanan and Shmatikov, 2010). Today's techniques can often enable data relating to search terms, websites visited, GPS positions, and IP address, to be linked back to an identifiable individual.

Consistent with the definition in the OECD Privacy Guidelines, this paper assumes a broad understanding of personal data, which transcends the various categorisations highlighted above.

Purpose and scope

Understanding the economic value of personal data within existing legal and regulatory frameworks is an extremely complex task. Using the tools and data that are currently available, findings are tentative at best. This paper focuses only on various methodologies for measuring the monetary value of personal data that emerges from market, not the full social and economic impacts.

Markets are the most efficient tools we currently have for setting prices for goods and services. But they are not perfect. The mechanisms of supply and demand determine prices in markets but there are other benefits or costs (known as "externalities" in economic literature) that may not be captured in the market price. For example, the value of a city park is not simply the price for which the land could be sold on a market, but also includes the benefits derived by visitors and the increased value of homes in the neighbourhood resulting from having a park close by. Using this analogy, this report only looks at the monetary value of personal data in markets (market value of the park) but not at the positive and negative externalities deriving from its use (analogous to increased property values in the neighbourhood resulting from the park, the enjoyment of park visitors, and the potential for delinquency in the park).

Externalities can be significant and should not be discounted in policy making. The value of personal data is intricately tied to its potential socio-economic impact. Many of the uses of data that create direct value don't necessarily involve a market transaction or can be measured by a market transaction—but the economic and social impact is direct. One example is the use of personal data to avoid duplicative testing/misdiagnosis, etc., in healthcare. This can create real and direct financial benefits, but the total benefits tend to go far beyond that. The *personal* value of using an electronic health record is improved treatment for the patient—and this undoubtedly has direct monetary value in the form of reduced costs, better outcomes, etc. But there are also valuable *socially beneficial* uses that create (or could create) value—e.g., for research into new drugs, identification of epidemiological trends, or improved medical protocols. Attempting to separate out these differences would help enable policy makers to be better informed when making rules governing the collection and use of personal data.

The ability to store data indefinitely, to combine it with other data, and apply analytics can greatly benefit both individuals and organisations. But using personal data in ways not anticipated when it was collected challenges core privacy principles. The insights enabled by analytics into an individual's trends, movements, interests, and activities, likewise bring risks of unfair discrimination. These and other privacy issues are not addressed in this paper. Nor does the paper explore the economic costs that irresponsible or illegal uses of personal data can impose on individuals, organisations and the societies in which they function.

Focusing only on methodologies of monetary valuation, therefore, does not capture the full social and economic benefits and costs of personal data but it does help lay an important foundation for future work. Any analysis that eventually seeks to understand the full social and economic impacts should begin with an analysis of markets where the good is exchanged.

Going forward, one approach to consider is to develop a conceptual framework to describe how personal data can create economic and social value, and identify potential indicators for those values. This would involve examining industry sectors that go beyond the usual advertising and social-media examples (e.g. healthcare, finance, automobile manufacturing, etc), and how insights gained from data analytics are already being leveraged to create value in each—some of it internal to the organisations that collect it to create new operational efficiencies; some through market exchanges of data; some via combining data from different sources; and so on. This approach could also potentially identify other indicators of the economic growth that can result from the value of personal data.

To re-iterate, the values emerging for personal data from the various methodologies are imprecise but they do take an initial step towards helping understand some broad ranges of market values for this increasingly important intangible asset.

The personal data economy: Mapping value chains and business models

A review of value chains can help identify the roles of different players and point to areas of market activity that enhance monetary values. Beginning with the value chain also helps provide a broad view of the activity and locate potential areas that policy makers may want to consider further.

This section explores personal data value chains, tracing the data lifecycle from capture to usage, and the business models employed for creating and realising the value of personal data along the way. The emphasis is on how personal data is enabling innovation in the form of new products and services, process efficiencies, new forms of analysis and the creation of value for both producers and users. These value chains are put into context by looking at specific ways personal data is used in various sectors of the economy. Application of the governing privacy frameworks may impact the value chain, and may vary in different countries and even within different sectors of the same country.

Personal data is collected in a variety of ways:

- Data can be *volunteered* or *surrendered* by individuals when they explicitly share information about themselves or about third parties (e.g., when someone creates a social network profile, enters credit card information for online purchases, provides his/her personal information as a condition of registration to a given on-line service, or posts information about a friend, colleague, family member, etc).

- Data can be legally *observed*, captured by recording the activities of users – in contrast to the data they volunteer (e.g., Internet browsing preferences, location data when using cellular mobile phones or telephone usage behaviour).

- Data can be *inferred*, based on the analysis of personal data (e.g., credit scores can be calculated based on a number of factors relevant to an individual's financial history). Personal data can be also *inferred* from several pieces of seemingly anonymous data (see Narayanan and Shmatikov, 2010)

Each type of personal data, volunteered, surrendered, observed or inferred, is initially collected or accessed, then stored, aggregated, processed and finally used and analysed. Each of these steps has some special features and could potentially involve different stakeholders (Figure 1). The personal data lifecycle can be presented as following a four-step[2] value chain:

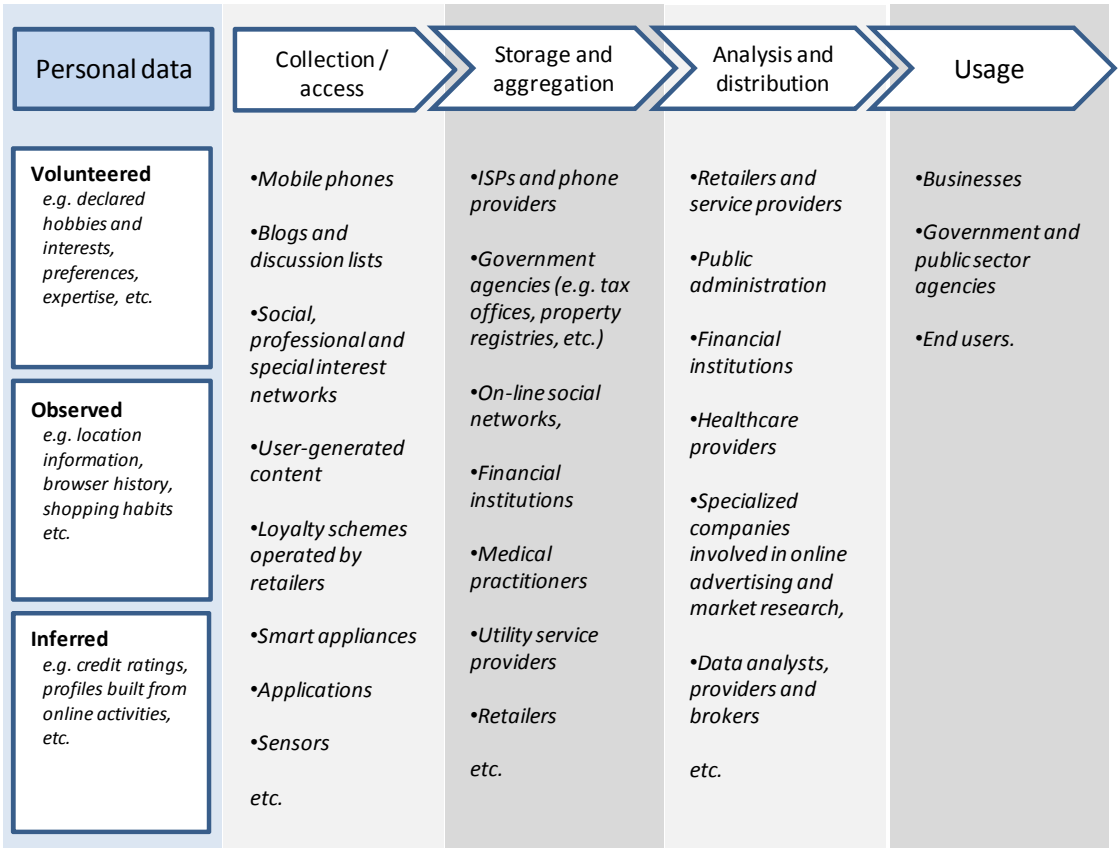
- Collection / access

- Storage and aggregation

- Analysis and distribution

- Usage

Figure 1. Personal data value chain



Source: OECD, based on WEF (2011).

A wide range of stakeholders are involved throughout the value chain, including individuals, businesses, public institutions, non-profit organisations, etc. Some stakeholders are involved only in selected parts of the value chain. For example data brokers do not typically *use* the personal data but rather process and sell it on. Other stakeholders can be involved in all the steps through the value chain. For example an airline or retailer can collect personal data through a customer-loyalty scheme, then store and aggregate them, and finally process and use them in its own business model (e.g., by marketing specific offers to targeted customers).

Finally, it is worth noting that the delineations between volunteered, surrendered, observed and inferred data are not clean. What is volunteered, surrendered and observed has a major impact on what is inferred, and what is volunteered and surrendered is often inaccurate.

Collection / access

The first step in the value chain is the collection of personal data or access to personal data within the applicable legal framework. This process covers all sectors of the economy and data are gathered from a myriad of sources. This collection or access step in the value chain has experienced a significant transformation with the growth of the Internet and other communication technologies that are generating large amounts of data. For example, mobile network operators have detailed data on individuals including their location (via mobile phone networks) and their call logs. Internet service providers potentially have access to very detailed information on the Internet use of their subscribers. Even in non-ICT sectors,

retailers now have much better tools to track sales to customers using loyalty cards and even in-store tracking technologies using sensors. This section will provide various examples regarding the collection of personal data.

Generally data can be collected directly or indirectly. Examples of direct data collection include the records of purchases and customer loyalty schemes operated by retailers, transport and hospitality services providers. Other data sources include customer identification, logon and transaction authentication operated by a range of phone, Internet and other services providers and user generated content in the form of e-mail, contributions to blogs and discussion lists, personal and professional websites, uploaded photos and videos, and participation in online social, professional and special interest networks.

Loyalty schemes operated by retailers and services providers enable them to build up a profile and get a greater understanding of their customers, decrease the transaction costs for both parties, and offer loyalty discounts and special offers. Logon and other forms of transactional logs give insights into user behaviour and needs, enabling service providers to offer enhanced services tailored more specifically to user needs. The wealth of user-generated content that has emerged over recent years also offers opportunities to develop profiles and target specific individuals with product and service offerings likely to be of most interest to them.

In some cases, data is gathered more indirectly, which is related to the emergence of newer business models. In these models data gathering can involve a range of activities such as recording the location of someone using a cellular mobile phone or registering and recording Internet location information for use in content and service delivery of relevance to users at a particular location or to limit access to content based on geographic boundaries. Other sources for observed personal data include browser history, page visits or downloads for tracking the online activities of users. This can also include purchases and transactions; access and use of specific apps; home and office automation, smart-grids and security (e.g., CCTV, sensors, etc.). It can also be used to help service providers identify potential security threats (e.g., where an account is accessed from an IP address in a different region or at a different time of day than is usual). As network technologies make the tracking of things and people across a wide range of activities feasible, new business models have emerged. These tend to focus on understanding individual consumers better in order to provide tailored products and services, to reduce consumer search and transaction costs and to increase the efficiency of suppliers and providers, be they private or public sector entities.

Finally, data can also be created through analysis rather than being captured, and can include the development of profiles and preferences from online activities that can be used in targeting advertising and offers; the establishment of credit ratings; and physical tracking and tracing through phone and/or Internet geo-location, sensors and home or office automation, smart-grids, etc. Personal data processing and analysis activities are discussed below.

Storage and aggregation

Once data has been collected, it can be stored and aggregated. This represents the second stage in the value chain. The individual data elements are organised and stored in datasets that can be used for further processing and analysis.

Numerous personal data records such as contact and credit or charge card details, account and login authentication details are stored by a wide range of services providers such as ISPs and phone providers, retailers; transportation firms, medical practitioners; utilities and government agencies. User generated and submitted content is also stored by a range of service and content providers, including social and professional networks, special interest social networks, blogs, photo and video sharing sites, and e-mail

providers. Because of the potential cost efficiencies, organisational and individual data are increasingly stored remotely and accessed online.

Data is also aggregated and stored by many actors in the value chain. Browser histories, Internet tracking and search records are kept by ISPs and search services providers; medical records are stored by medical practitioners and a range of public and private sector agencies, healthcare insurers, employers and others; financial records are stored by retailers, banks and other financial institutions, employers and taxation agencies; and locational information is stored by mobile phone operators, ISPs, utilities, transport operators and others.

Analysis and distribution

The third step in the personal data value chain takes the collected and stored data and combines it with other information to develop detailed profiles, records, and macro trends that can be used for various purposes. One key element of the value added in this step is the merging of data from different sources into profiles and using analytics to infer information that may not be otherwise available. These sources may include data sets that are publicly available, proprietary data owned by businesses, and data from institutional research. The possibility of such aggregation present unprecedented opportunities for drawing unique insights, creating tremendous opportunities for new products and services. It is worth noting that data can go through several rounds of analysis and distribution, with additional data added with each iteration. When the insights are used to establish a more refined personal profile, data analytics firms often resell the combined profiles in the market and thus are an important source of data on the monetary value of personal data. Examples of other types of insights drawn from such analyses include improved customer service and product quality, drug interaction issues, and common daily traffic patterns.

This work is often done by firms with developed infrastructure, strong analytical skills and developed distribution networks. This can involve both more traditional players in the value chain and new firms that have emerged in response to new needs and opportunities. Among the more traditional players in personal data processing are retailers and service providers operating customer relationship management (CRM) software, business intelligence systems and loyalty programmes. More recent players include those involved in online advertising, market research, specialist data analysts, providers and brokers. Indeed, a "tracking industry" has emerged that is driving innovation in advertising (See Box 1).

Box 1. Personal data analytics, distribution and usage: The case of online advertising

Online advertising involves a number of players, including web publishers, advertisers and advertising network intermediaries who connect web publishers with advertisers seeking to reach an online audience. Advertising networks play a central role. An "ad network" is a company that sells advertisements on behalf of website publishers. Often referred to as "third-party ad networks" because they serve a broad range of publisher partners, advertising networks purchase available advertising space from publishers and then resell it to the ultimate advertisers. The relationship is beneficial for both parties. Web publishers benefit from advertising as it enables them to derive revenue from their content without having to charge subscription or other user fees, while advertisers reach their target audience. These intermediaries are particularly important for small web publishers who cannot afford a large advertising sales force or the search costs associated with finding potential advertisers. Advertisers need advertising networks to promote their products effectively to relevant audiences without the significant search costs of locating and negotiating directly with individual publishers.

Advertising networks use contextual, vertical and behavioural strategies for matching advertisers with users of Internet content and services. Contextual networks allow the advertisers to bid on keywords on publisher websites within the network's inventory; vertical networks typically group together similar publishers within their inventory and offer them to advertisers (e.g. automobile companies are likely to want to advertise on sites geared toward viewers interested in cars); and behavioural networks employ behavioural targeting to direct specific advertisements to certain viewers by collecting and using data based on user browsing behaviour across multiple websites in order to categorise likely consumer interest segments for use in targeting.

Advertising networks are operated by a range of players. In the early days they were typically either a specialised part of major advertising agencies (e.g. WPP Group, Omnicom, Publicis, etc.) or operated by the major portals (e.g., Yahoo!, MSN, etc.). Since the mid-2000s specialist advertising network firms have emerged and, with the growth and potential of behavioural targeting, major Internet firms with access to rich user information have moved into advertising network operations, often through acquisition (e.g., AOL through Advertising.com and TACODA; Google through DoubleClick; Yahoo! through Right Media and Blue Lithium, etc.). Increasingly, large advertising networks include a mixture of search engines, media companies and technology vendors.

These networks are supported by specialist advertising and data exchanges. An "ad exchange" is an auction-based marketplace where advertisers can bid to place advertisements in the space offered by websites. A "data exchange" is a marketplace where advertisers bid for access to data about customers. The data can be that collected through the tracking and tracing of users' online activities and/or from offline sources (e.g. national statistics, census data, etc.). Increasingly, data are analysed and combined, and a user's profile developed by specialist data analysers.

Writing in the Wall Street Journal, Angwin and McGinty (2010) described a "tracking ecosystem" showing how, when a user visits a website, tiny tracking files ("cookies") watch what they do and develop a profile of the user's behaviour. Often, a tracking company will sell this information directly to advertisers, but some sell the data on a data exchange, which can combine it with other sources of off-line personal data (e.g., census data, real estate records, vehicle registration, etc.). This enhanced data, typically in the form of a user 'profile' is then sold to advertisers looking for consumers matching that profile. Linking this to the unique identification code embedded in cookies on the user's computer or handheld device, an advertiser can target advertisements at the particular user, and/or simply buy advertising space on websites matching the interests of that user profile through an advertising exchange. Thus delivering advertisements to users who are likely to be interested in them.

Behavioural targeting is effective for advertisers because it produces more 'hits'. Surveying nine of the top 15 advertising networks, Beales (2010) found that behaviourally-targeted advertising accounted for around 18% of total advertising revenue during 2009 (USD 595 million), cost 2.68 times as much as run-of-network advertising and was more than twice as effective at converting users who click on the advertisements into buyers – a 6.8% conversion versus the 2.8% conversion from run-of-network advertisements. It also provides product and service information that is likely to be of interest to the particular consumer, thus reducing search and transaction costs.

An interesting and relatively new platform for both data collection and targeted advertising is mobile phone operating systems. Two smart-phone operating systems dominate (Apple's iOS and Google's

Android) and each of the backers also has business interests in targeted advertising (Box 2). There is a growing industry in assembling data collected from mobile phones into profiles of cellular mobile and smart phone users. Mobclix, an advertising exchange, matches more than 25 advertising networks with some 15 000 apps seeking advertisers. By tracking a phone's location, Mobclix also makes a "best guess" of where a person lives. Mobclix then matches that location with off-line spending and demographic data from Nielsen (Thurm and Yukari 2010). This increases the effectiveness of mobile advertising and enhances the value of the information for users by making it locationally relevant.

Box 2. Platforms for smart phones

With the rapid increase in use of smart mobile devices, making information and advertising locationally relevant is a major area of innovation. The two most popular platforms for new smart phones are Apple's iOS and Google's Android, and Google and Apple run the two biggest services, by revenue, for putting advertisements on mobile phones.

Google was the biggest recipient of data from smart phone apps in test run by the Wall Street Journal (Thurm and Yukari, 2010). Its AdMob, AdSense, Analytics and DoubleClick units received information from 38 of the 101 apps tested. Google, whose advertising units operate on both iPhones and Android phones, says it does not mix data received by these units. Google's main mobile advertising network is AdMob, which it bought in 2010 for USD 750 million. AdMob lets advertisers target phone users by location, type of device and demographic data, including gender or age group.

Apple operates its iAd network only on the iPhone. Eighteen of the 51 iPhone tested apps sent information to Apple. Apple targets advertisements to phone users based on what it knows about them through its App Store and iTunes music service. The targeting criteria can include the types of songs, videos and apps a person downloads.

The distribution of the processed information takes many forms. For major players, a good deal of the data collected and processed is used internally in CRM software, through loyalty programmes and transaction records to tailor product and service offerings to customers. Others trade and transfer the data they collect and some purchase processed data as a service (e.g. credit ratings, profiles, etc.) from data analysts and brokers and/or on data exchanges (see Box 2).

In the context of online advertising, a **data exchange** is a marketplace where advertisers bid for access to data about customers.[3] The data can be collected through tracking and tracing of users' online activities and/or from offline sources (e.g. national statistics, census data, etc.). For example, with third party tracking, the first time a website is visited, the data collector installs a tracking cookie, which assigns the computer a unique identification number. Whenever the user visits another website affiliated with the tracking company, it can note where the user has been over time, thereby building up a more complete profile of the user's tastes and habits. Within seconds of visiting such a website, information detailing a Web surfer's activity may be auctioned on a data exchange, such as that run by BlueKai (Angwin, 2010). According to their website, the BlueKai Exchange is the world's largest data marketplace, with data on more than 300 million users offering more than 30 000 data attributes; it processes more than 750 million data events and transacts over 75 million auctions for personal information a day.[4]

Box 3. Data brokers

Data brokers are firms that gather and merge aggregated information on individuals that is then sold for various uses such as employment background checks, the issuing of credit and for law enforcement purposes. The data is typically collected from public records or various sources where individuals have provided information. Since their business models are usually based on selling data about individuals, the value the firms and customers are assigning to the data in the transaction can contribute to an understanding of the value of personal data.

Specific profiles of data brokers range from specialised business-to-business companies, through regular data brokers, to simple localisation services. The specialised business-to-business companies (e.g. LexisNexis) offer a complete complex background check of all possible information for businesses about potential business partners. Their scope is much broader than personal data only. Regular data brokers (e.g. Intelius, Locate Plus), provide information solutions for consumers and small businesses using public records and publicly available information. Their products help people find other people, verify the identities of individuals they encounter, manage risk, ensure personal safety, etc. Finally localisation services (e.g. LocatePeople.org, MelissaData.com, 123people.com) provide personal addresses of individuals to data marketers, or offer simple services used just to localise people, their phone numbers, e-mail addresses, etc.

It appears that the number of firms offering these services has been growing but it is difficult to present reliable statistics since there is no standardised classification of data brokers. Some belong to the business classification category "Hosting, and Related Services" and others are classified as Credit Bureaus.

At the time of writing, privacyrights.org listed 180 US-based online information brokers. They include companies like LexisNexis, which claims to conduct more than 12 million background checks a year, Acxiom, with annual revenues of around USD 1.2 billion, Experian, which focuses on credit information, and many smaller operations. Similarly, companies such as Everify.com offer instant background checks that include detailed information, such as the full name and aliases, birth date, up to 40-year address history, current and previous phone contacts, associates, relatives, neighbours, criminal records, misdemeanors, felonies, sex offender check, lawsuits, judgments, properties, marriage and divorce records, email addresses, and social network profiles for USD 19.95. The LexisNexis service Accurint provides a detailed, itemised price list that demonstrates for how little many individual items of personal data are traded.

Usage

Once the data have been collected, stored and analysed, they are often made available to end users in markets, representing the end of the value chain. The end users generally purchase profiles of individuals (or firms) in order to supplement their own business activities. If these transactions can be evaluated, they can provide into how highly the market values the personal data.

Personal data are used in many ways by businesses, public sector agencies and end users. Categorising the different uses of personal data is beyond the scope of this report but there are a number of broad patterns emerging. Personal data are typically used to better serve customers and improve the efficiency of transactions between the entity and end-users. Personal data are also being aggregated to improve business operations efficiencies, as well as to identify macro trends in a number of different sectors, including healthcare, transportation, and safety.

Businesses typically use the information to better understand their customers, and in many cases, offer them targeted advertisements or services. There has been a shift in recent years from Internet business models based on *Free: The future of a radical price* (Anderson, 2009) towards *Not for Free: Revenue strategies for a new world* (Berman, 2011). The former model is typically supported by advertising (Box 1). In the latter, Berman (2011) argues that media and entertainment industries and content publishers need a greater variety of revenue models than the commonly employed free online/paid offline approach or pure advertising support. Berman *et al*. (2010) noted how online advertising-supported business models have yielded a lower unit return for traditional content publishers – with a print newspaper reader bringing in 18

times the value of an online reader, and a broadcast television viewer three times the value of an online viewer. They attribute the value discrepancy to downward pricing pressure due to a surplus of online inventory and reliance on advertising networks, and online video supporting far fewer advertisements than broadcast television (Berman, et al., 2010).

One proposed solution is to move to a consumer-centric model, using data to provide more value to the consumer. This involves creating incentives to gain consumer insights and then offering value to the consumer in exchange (Berman, et al., 2010). This study suggests that media and entertainment firms need to find innovative ways to "upgrade" traditional sponsored revenue models so they appeal to the digital consumer. In particular, it is important that they move away from the "one-to-many" approach of traditional media and leverage the advantage of digital platforms to target and interact with consumers in a segmented or individualised way. They need to go 'beyond advertising' and focus on consumer-centric marketing tactics.

Governments and public sector agencies: Governments increasingly rely on personal data, obtained not only from third parties but also directly from individuals, to administer various programmes. Examples include social service programmes, administering tax programmes or issuing licenses. Data is also commonly used to support hundreds of regulatory regimes ranging from voter registration and political campaign contributor disclosures to employee identity verification and child support obligation enforcement. Other uses can be to maintain vital records about major lifecycle events, including birth, marriage, divorce, adoption, and death; or to operate facilities such as toll roads and national parks. The increased reliance on personal data helps to provide services to a larger population, diminishes the perceived inequality of subjective determinations, reduces the costs of litigating decisions and maintaining more skilled personnel, and enhances accountability (Cate, 2008). What has changed is the increasingly extensive use of data analysis and data matching in pursuit of compliance, and a result of this is that personal data are being used in new ways.[5]

Several of the uses detailed above are of value to individual *end users*. Customer-centric business models use personal data and behavioural targeting to enhance services. Targeted advertising reduces search and transaction costs enabling people to more easily find products and services of interest to them and avoid the inconvenience of seeing advertising material of no interest to them. More cost-effective advertising also better supports the free-to-use advertising-supported services that are widely used (*e.g.,* search services such as Bing). Social networks also provide value to end users, not only in their social lives but also in professional networks (*e.g.,* LinkedIn) and the opportunities that arise through such networks. These might include introductions to collaborative and business opportunities, and prospective employers checking applicants.

Special interest social networks can also be of enormous value to end users. An illustrative example is the network PatientsLikeMe (Box 4) that not only enables people with a medical condition to interact with, derive comfort and learn from other people with the same condition, but also provides an evidence base of personal data for analysis and a platform for linking patients with clinical trials. The business model depends on aligning patient interests with industry interests, and PatientsLikeMe sells aggregated, de-identified data to its partners, including pharmaceutical companies and medical device makers, to help them better understand the real-world experiences of patients as well as the real-world course of disease. Apart from that, PatientsLikeMe also shares patient data with research collaborators around the world. But the PatientsLikeMe example also highlights one of the challenges in capturing the economic value of personal data. There is a monetary value attached to the data that is sold to partners but the service is free to end-users, many of whom place a high value on interactions with other within the community of people with similar challenges. The value to the end users is extremely difficult to capture and estimate when the services are provided without payment.

Box 4. Use of personal health records

The platform PatientsLikeMe is an example of a platform to enhance research in the healthcare area. It is an Internet-based community of patients diagnosed with life-changing diseases, doctors, and organisations. It enables users to share information that can be used for research on a particular disease and that improve patients' lives. In particular, the Internet platform permits the collection and sharing of real world, outcome-based patient data. Relevant data are transmitted to doctors, pharmaceutical and medical device companies, research units, etc. who have established data-sharing partnerships with PatientsLikeMe. Since most of the research in modern medicine relies on large amounts of data, this platform becomes an extremely efficient tool for healthcare professionals and industry organisations that are trying to treat the disease and, at the end of the day, for patients themselves.

Source : www.patientslikeme.com/

Methodologies for estimating the value of personal data

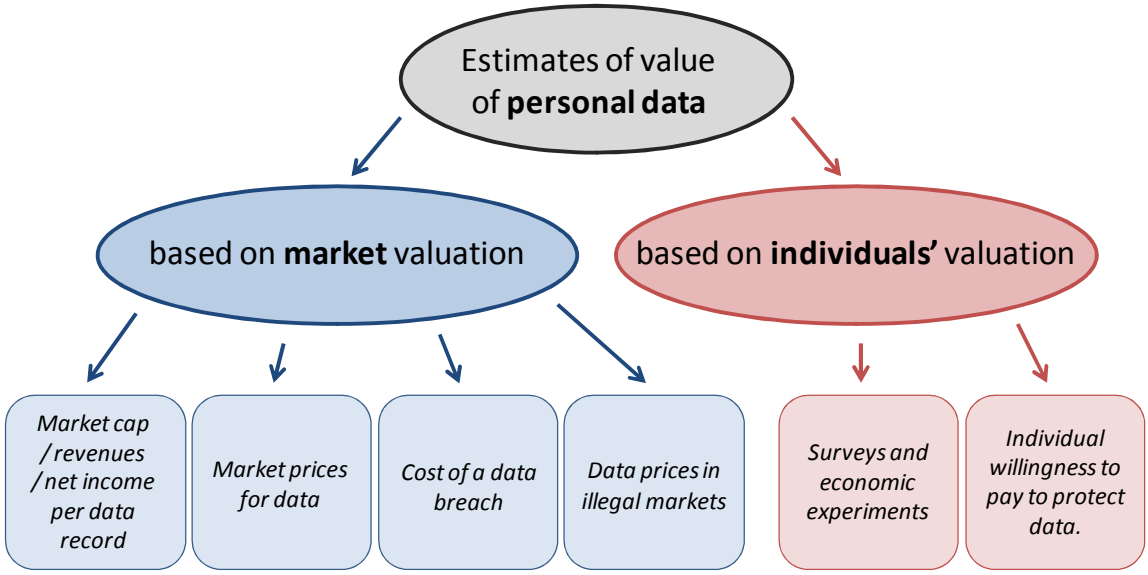
This section evaluates potential methodologies for determining the monetary value of personal data. As mentioned earlier, these methodologies do not capture the social and economic benefits of personal data, but rather the prices that markets are assigning to the data in different contexts. The section looks at six potential methodologies for developing preliminary estimates of monetary value and then describes benefits and drawbacks of each of the approaches.

Most of the approaches focus on the value assigned to an individual record or user. A record could be a single piece of information for one such as the person's age, but it could also represent an entire profile with demographic, economic and educational information in another context. At this stage it is not possible to develop even a basic definition of what a record would include. Instead, these analyses look at the value per record that can only be considered within the context of that particular market.

The methods are rudimentary and each measures different aspects of monetary value so they are not directly comparable. Yet, these monetary transactions and market valuations can still serve as an initial base for expanding our overall understanding of personal data in the economy.

There is no commonly accepted methodology for estimating the value of personal data. Possible approaches to do so rely either: *i*) on market valuations of personal data, or other related market measures, or *ii*) on individual perceptions of value of personal data and privacy (see Figure 2).

Figure 2. Estimates of value of personal data



Source: OECD

Measures that are based on **market** valuation refer to values of personal data records that can be observed or derived from the markets. Observable measurable proxies of value of personal data include *i*) market cap/revenues/net income per data record *ii*) market prices for data *iii*) cost of a data breach; and *iv*) data prices in illegal markets.

Measures based on **individuals'** valuation include results from surveys and economic experiments, as well as data on willingness of users to pay to protect their data.

There are two important points that should be kept in mind in this context. Firstly, it must be highlighted that there is no single, perfect measure of value of personal data. Each of the presented measures might suffer from certain methodological biases and measurement errors. It should also be re-emphasised that it is crucial not to isolate personal data from the underlying context of the business model in question.

Secondly, most of the available statistics on personal data refers to the US market. Therefore, illustrations that complement the presented measures, in most cases, present the US estimates. It should be emphasised that a potential extension of these illustrations onto other markets is not straightforward, and should take into account numerous important regional factors, for example cross-country differences in privacy legislations.

Table 1 summarises the discussed measures.

Table 1. Summary of measures of value of personal data

Indicator	Description	Benefits	Potential Drawbacks
Indicators based on market valuation			
Financial results per data record	Aggregated market cap (revenues, or net income) of a company divided by the total number of personal data records used by this company.	- Relatively easy to identify. - Reflects actual economic value added generated through personal data.	- Likely to be inaccurate, as numerous other components impact market cap / revenues / income of a company. - Possible synergy effects could lead to overestimates for firms with larger datasets. Appropriateness of this approach depends on what portion of turnover is directly tied to personal data.
Market prices for data	Price per personal data entry offered on the market by data brokers.	- Relatively easy to identify; - Reflects market value of a given, specific data entry.	- Apart from the data value, it includes the cost of data search and processing. It also neglects the context in which the data is sold, which has a large influence on the demand (and price) for data.
Cost of a data breach	Economic cost of a data breach (for firms and individuals) per data entry.	- Reflects a real market value and a portion of the risk that companies must protect against.	- Captures market costs of damage caused by data breach rather than value of data themselves. Does not include the costs of damage to a firm's reputation.
Data prices in illegal markets	Estimation of prices of personal data (per data entry) in illegal markets.	- Reflects market value of a given, specific data entry.	- Difficult to measure and only applies to the context where the data is used again to obtain other benefits illegally. Because criminals must balance the risk of detection and punishment, the value of the personal data is likely undervalued by such an approach.
Indicators based on individual (data subjects') valuation			
Surveys and economic experiments	Valuation of personal data in monetary terms are reported / revealed by individuals in surveys / economic experiments.	- No ambiguity in data identification. - Captures the pure economic value of personal data from an individual perspective. - Results usually can be used for comparative studies (across economies and across various types of data).	- Hypothetical value not verified by the market. Previous research shows that a person's valuation of their own personal data is highly sensitive to context, meaning that the way questions are phrased could significantly alter the responses.
Individual willingness to pay to protect data.	Amounts that individuals are ready to spend to protect their personal data.	- Captures the pure economic value of privacy from an individual perspective.	- Captures individually perceived aggregate costs of damage caused by data breach, rather than value of data themselves.

Financial results per data record

The first method for estimating the value of personal data looks at the financial results of a company such as market capitalisation, revenues and net income on a per-user or per-record basis as a way to capture the market values of the data in each record. This type of analysis only works for firms that either derive most, if not all, of their revenues from personal data or firms that separate out revenues for personal data activities. Many companies that work with personal data are not publically traded or do not report information in sufficient detail to perform these analyses.

Each of the financial indicators measures something different. The market capitalisation approach is an approximation of the market value of the company (assigned by the market valuation of the stock price) averaged across the number of records the firm holds. In theory, the market capitalisation reflects the value of the firm with all future earnings considered and discounted to their present value. In reality, the market values of firms often fluctuate with general market sentiment and other economic shocks that may not be tied to the underlying value of the personal data.

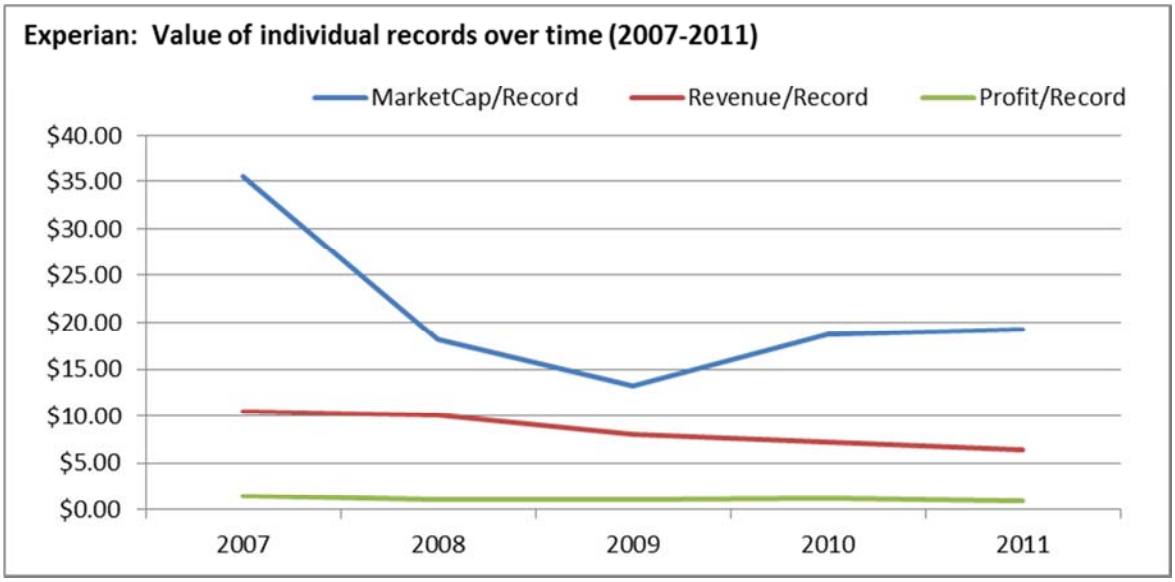
The second indicator looks at the reported revenues of the firm averaged across the total number of records. It is a way to capture the "productivity" of a record, essentially the amount of revenues it brings in each year. The measurement may be considered a more realistic measure of the monetary value of data than market capitalisation because it reflects the average amount of revenue that can be attributed to each data record and is less likely to be buffeted by changes in market sentiment and general economic shocks.

The final indicator is profits per data record and represents the net profit (revenues minus costs) per data record held by the firm. Revenues per record arguably remains the most robust indicator derived from financial information because it captures how much the firm earned (in total) for the data while profits per record would include costs that would fluctuate and may not be connected to the underlying value of the data.

For illustration, the data broker Experian reported 2011 revenues of USD 4.2 billion and had a market capitalisation of around USD 10 to 12 billion from a reported 600 million individual and 60 million business credit application and payment history reports. This works out to a capitalisation of around USD 19 per record and annual revenue of around USD 6 per record. Profit per record was considerably lower in 2011 at roughly USD 1 per record (Figure 3).

Figure 3. Experian: Key financial information per record held

Individual and business reports, millions (2007-2011)



Source: OECD based on Experian annual reports 2007-2011

An examination of Figure 3 shows that Experian's market capitalisation per record has fluctuated in a pattern that is similar to overall market trends between 2007 and 2011. The market capitalisation per record declined sharply at the beginning of the economy crisis before reaching a bottom in 2009. By 2011, the value was beginning to increase again. By contrast, the measure of revenues per record had a more gradual

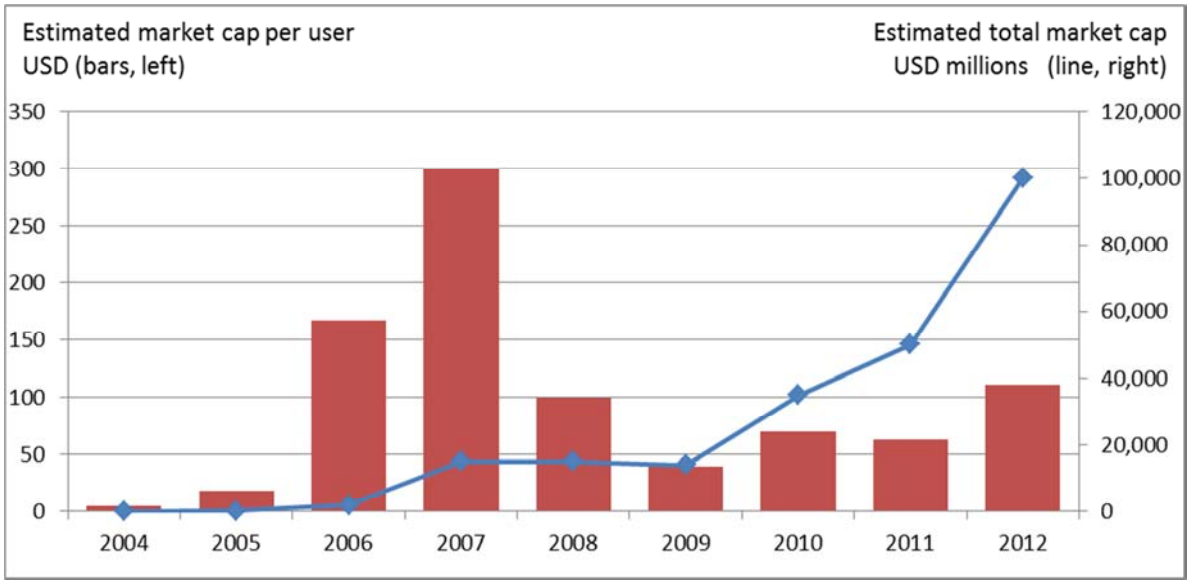
and stable decline between 2007 and 2011. In 2007, the annual revenues per record were USD 10.55 but declined to USD 6.42 over the period. Profits per record also declined but by a smaller percentage. In 2007, Experian's average annual profit per record was USD 1.40 but declined to USD 0.94 in 2011.

It can be illustrative to compare the financial indicators per record from a data broker such as Experian with another company that derives nearly all of its income from advertising linked to personal data, Facebook. In May 2012, Facebook prepared an initial public offering of its stock which put the valuation of the company at just over USD 100 billion for a user base of roughly 900 million. This works out to a valuation that is the equivalent of USD 111 per registered user.

Figure 4 shows that the value of an individual record (user) has fluctuated significantly between 2004 and 2012 based on estimates of the market value of Facebook. It is important to note that since Facebook has only recently been traded publically, the bulk of the valuations are based on the size of private investments into the company for a certain percentage of equity. Even with significant volatility, the valuation per user has remained between USD 40 and USD 300 since 2006 at the same time the firm's market capitalisation rose from USD 2 billion to USD 100 billion.

Figure 4. Facebook estimated market capitalisation per user

2004-2012

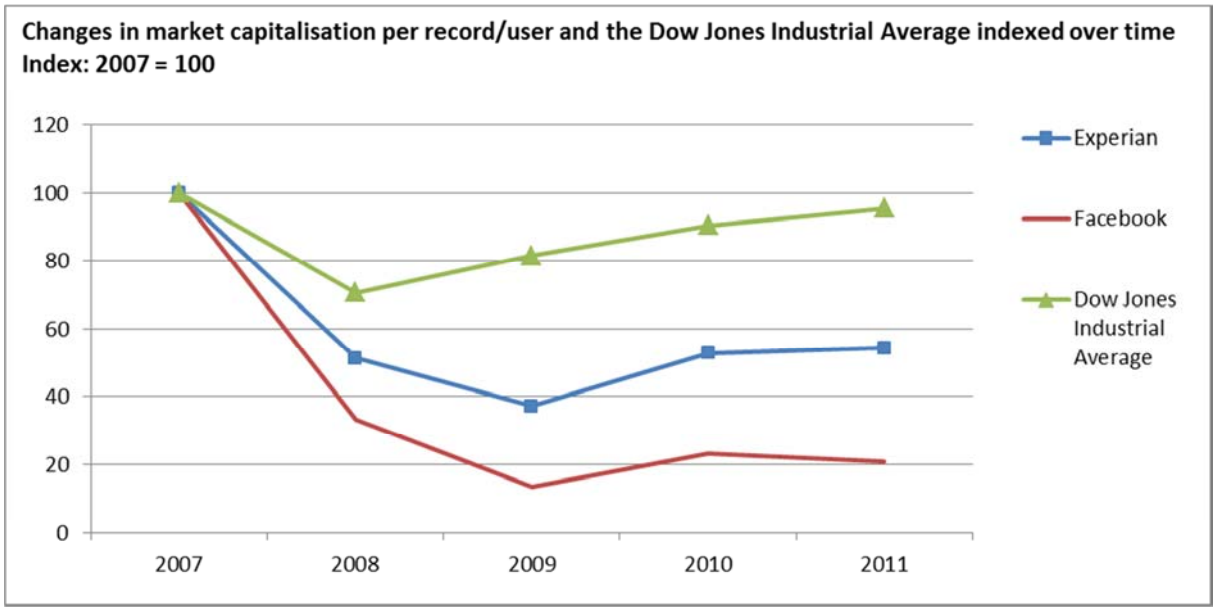


Source: OECD graphic derived from TechCrunch and CBS
http://techcrunch.com/2011/01/10/facebook-5/
www.cbsnews.com/8301-505250_162-57370133/number-of-active-users-at-facebook-over-the-years/

The data broker Experian experienced a similar change in valuation per data record held between 2007 and 2011. Figure 5 presents a normalisation of the valuations per record or user for Experian and Facebook to show changes over time when compared with the Dow Jones Industrial Average (DJIA) which tracks the broader stock market in the United States. In terms of valuations per record or user, Facebook and Experian follow a similar path but trail the overall recovery shown in the United States by the DJIA.

Figure 5. Changes in market capitalisation per record/user and the Dow Jones Industrial Average over time

Index: 2007 = 100

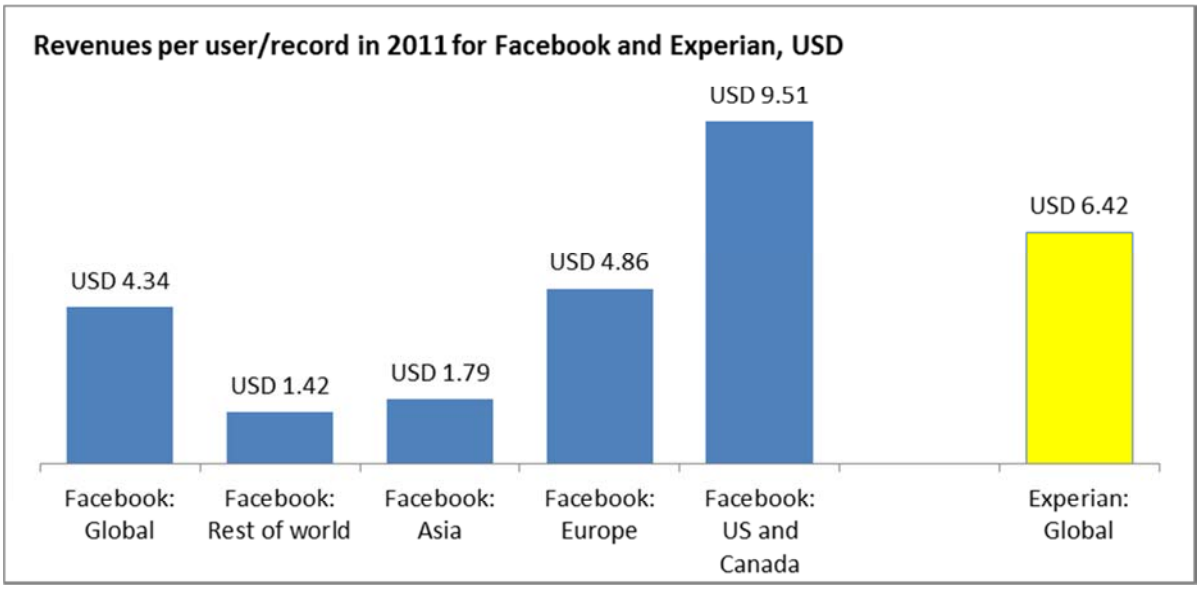


The market capitalisation figures from Experian and those imputed for Facebook seem to lend support to the notion that valuations of personal data based on market capitalisation will be affected by overall market sentiment and outside shocks that may be unrelated to the underlying value of the data in the records.

Revenues per record or user provide a better gauge of the monetary value because they are directly tied to the amounts that others are paying for access to the data. Experian and Facebook employ very different business models, store different types of data and use the data they collect differently. It is interesting then that both firms reported similar levels of revenue per user or record, despite these differences (Figure 6).

Figure 6. Annual revenues per user/record, Facebook and Experian

2011



Note: Facebook figures are in terms of advertising revenue per user and are from TechCrunch: http://techcrunch.com/2012/05/03/stats-facebook-made-9-51-in-ad-revenue-per-user-last-year-in-the-u-s-and-canada/. Experian figures are from their 2011 Annual Report.

Benefits of this approach

Given that publically traded firms disclose their financial statements and venture capital investments are often announced, the identification of these measures are relatively straight forward. Reporting requirements for public firms mean that data on market capitalisation, revenues and net income are available for a large number of firms. These data can be combined with the volume of personal data used as input of any given firm as a way to yield an illustrative measure of market value of a single personal data record.

Moreover, financial results are generated in a market environment, hence this measure reflects the economic value added generated through the actual usage of personal data in a market context. In other words, the net profit, revenues or market capitalisation are generated in the market, as the final outcome of the market forces. Consequently, a measure of value of personal data that relies on these financial results can be considered as a derivative of its true market valuation.

Drawbacks of this approach

There are several limitations related to the use of financial results (e.g., market cap, revenues or net income) divided by the number of users or records as a measure for value of personal data. There are numerous components other than personal data that impact financial results of a company, such as its human and physical capital stock, volume of other intangible assets and expertise. Consequently this measure could significantly overestimate the value of personal data, particularly in industries where personal data are not the main input factor of production.

The choice of financial measures used to value personal data is important. The measure of revenues per personal datum/user does have limitations because revenues only contribute to economic growth

insofar as they generate added value (or surplus). Net income per personal datum/user could be a more effective measure because it captures the value added by the firm using the data more precisely. Market capitalisation per personal datum may be the least reliable measure of the value of personal data, because for the majority of companies it is clouded by the many other items in the corporate value chain.

In addition, the relationship between the volume of personal data and financial outcomes may not be linear, even for companies that rely heavily on personal data. For example, it is likely that use of personal data in an economic activity could include synergistic effects, i.e., when for a given company the economic value of a single isolated data record is smaller than the economic value of the same data record in a large, consistent dataset. Such synergies could in turn lead to significant estimation biases.

Finally, the financial results of a company can also depend largely on market trends, random shocks and speculation. This means that measures of personal data could be potentially imprecise and fluctuate over time following general market trends or speculative activity, rather than the intrinsic value of the data.

Market prices for data

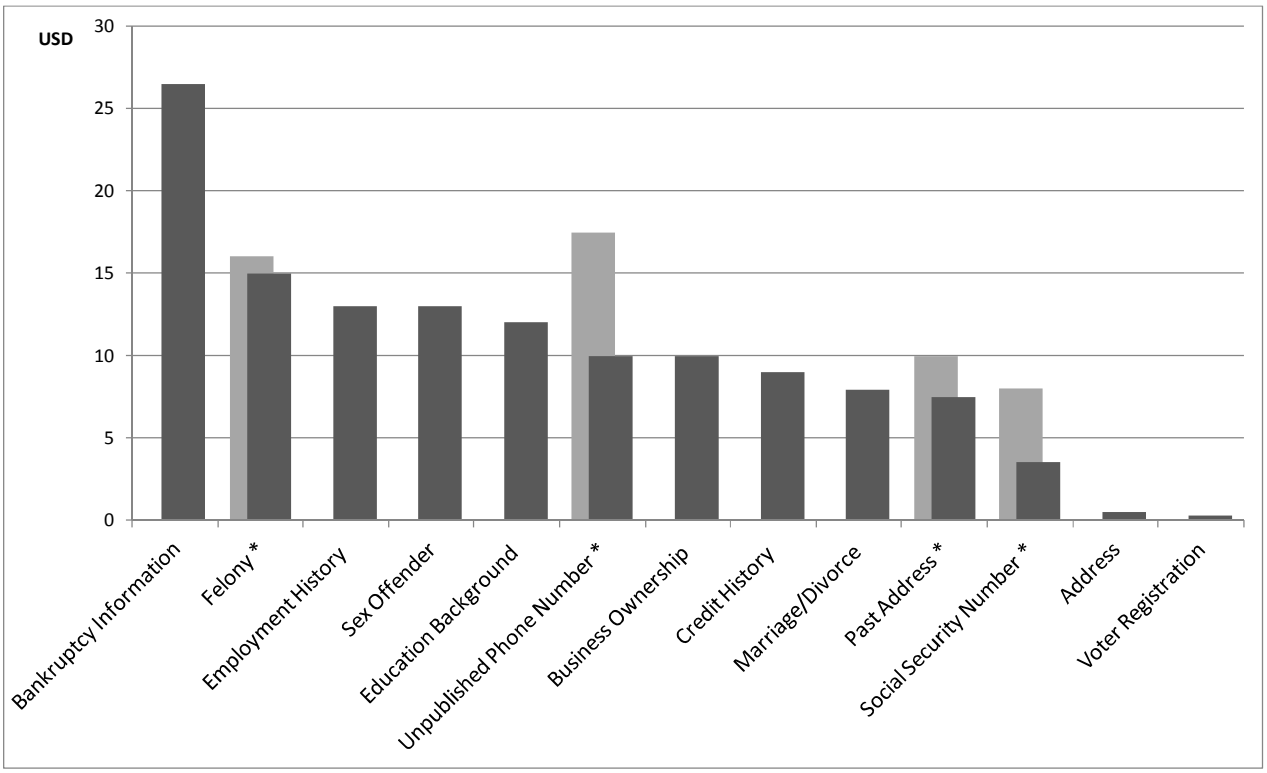
One of the most important ways to understand the monetary value of data is to look at its price in competitive markets. The price for data in the market is determined by the intersection of supply and demand. But personal data is a "non-rival" or "non-substractable" good in economics, meaning that the use of the data by one person does not diminish the stock of the good; the same record can be sold many times to many customers, and the same record can be used multiple times by the same customer. As a result, the market price for a record sold to one customer does not reflect the full monetary value of the underlying data but rather provides an indication of the market clearing price that individual customers pay for a copy of the data. The average revenue per record highlighted in the first methodology is comparable to the sum of the prices that all customers paid for an individual record over the course of one year.

Personal data transactions were typically handled between companies privately in the past, but the Internet is opening new markets and increasing the availability of price information. The prices for certain types of personal data are advertised on the Internet (e.g. personal background checks) while determining other prices requires a discussion with a sales representative and a quotation based on the type of information required. An illustrative example is the US market where numerous companies, called *data brokers* engage in these types of transactions selling personal data.[6]

One common feature of data brokers is that they engage in transactions using the data from third parties. Hence, prices demanded by these companies are an interesting indication of the market value of a personal data record. Figure 7 summarises some estimates that are derived from various online data warehouses (e.g. Aristotle, LexisNexis, DocuSearch, Experian, Merlin Data, Pallorium). At the time of writing, the following elements of personal data were available for various prices: USD 0.50 cents for an address, USD 2 for a date of birth, USD 8 for a social security number (government ID number), USD 3 for a driver's license number and USD 35 for a military record. A combination of address, date of birth, social security number, credit record and military is estimated to cost USD 55. These are only estimates, but provide some insight into the relative market values of difference pieces of personal data.

Figure 7. Market prices for personal data by type

(per record)



* two different prices provided by different providers.

Sources: Locate Plus (address Unpublished Phone Number Felony) Pallorium (address, past address Unpublished Phone Number Social Security Number) KnowX via Swipe Toolkit (Past address, Marriage/Divorce Bankruptcy Information Business Ownership) LexisNexis via Swipe Toolkit (Education Background Employment History Social Security Number Felony sex offender) Experian (Credit History) Voters online.com (voter registration)

Benefits of this approach

Market prices of personal data records have two main benefits that make them a useful measure to express the economic value of personal data.

First, similarly to the financial outcome per data record, market prices for data are relatively easy to interpret when the data is available. Several data brokers publicly announce their asking prices for various personal data records. These prices can be easily aggregated and compared across various brokers to become the proxy for market value for a single personal data record.

Second, just like many other companies in other sectors, data brokers operate within competitive markets. Hence, prices for personal data asked by data brokers are market prices that result from the intersection of demand and supply. Consequently, these prices at least partially reflect the real market price for obtaining a specific data entry within the market.

Drawbacks of this approach

There are a number of drawbacks in using market prices from various sources as a measure of the monetary value of the underlying data. The prices in the market also reflect search and other costs incurred by the data broker (such as the labour cost) and the possible search efforts saved by the purchaser. The data

may be available from a public source but require a considerable amount of effort to convert them into something that is easy to use for customers.

The price at which data is exchanged in an open market relates only to a specific context. The monetary value of a phone number, in and of itself, may be relatively low as a standalone item. However, when that phone number is tied with an income level and a set of particular interests, the monetary value of the phone number would likely increase for some market participants. So prices observed in the market need to be considered in context.

Finally, the quality of data provided by data brokers cannot be verified *ex-ante*, and may be flawed, unreliable or inaccurate. This risk is likely to be incorporated into the price of data, and additionally biases its true value.

Illegal markets

An approach that is somewhat related to the market price approach discussed above is the valuation of personal data in illegal (cyber-crime) markets. These are markets run as online forums (mostly within channels on Internet Relay Chat (IRC) servers), where cyber criminals buy and sell software, information and services as diverse as malware code, botnets, denial of service (DoS) attacks, scam page hosting, and last but not least personal data such as "government-issued identification numbers, credit card numbers, user accounts, email address lists, and bank accounts" (see Symantec, 2010; Panda Security, 2010). Some security firms such as Panda Security (2010) have identified and observed these markets as well as the offers posted in these markets.

Based on these observations, a range of magnitudes of the valuation of personal data by cyber criminals can be derived. According to Symantec (2010), for example, credit card information was the good most frequently advertised for sale in cybercrime markets in 2009, with 19% of all advertised items accounting for credit card information (see Table 2.).[7] The prices offered ranged between USD 1 and USD 30 in 2009. According to Symantec, the value would depend on several factors including for example demographics of the data subject and the intensity of its online interactions in particular in the context of e-commerce and e-banking.

But more importantly, the price varies over time depending on the total number of offers for credit card information available in the cyber-crime markets. This in turn depends on the amount of personal data stolen during data breaches (see Box 5). Credit card information is among the most commonly stolen and resold. In 2008, almost a third of all items offered in cybercrime markets were credit card details but this fluctuates from year to year depending on the amount of data breaches. There were more data breaches and consequently more credit card details for sale in 2008 than in either 2007 or 2009, according to Symantec (2010). As the number of breaches fell and the number of available card numbers declined in 2009, the maximum price asked for credit card information increased by 16% in 2009 to more than USD 30.

Table 2. Goods and services advertised for sale on cyber-crime markets

Overall Rank 2009	Item	2009	2008	Range of Prices in USD
1	Credit card information	19%	32%	0.85 - 30
2	Bank account credentials	19%	19%	15 - 850
3	Email accounts	7%	5%	1 - 20
4	Email addresses	7%	5%	1.70/MB - 15/MB
5	Shell scripts	6%	3%	2 - 5
6	Full identities	5%	4%	0.70 - 20
7	Credit card dumps	5%	2%	4 - 150
8	Mailers	4%	3%	4 - 10
9	Cash-out services	4%	3%	0 - 600 plus 50%–60%
10	Website administration credentials	4%	3%	2 - 30

Source: Symantec (2010)

One key difference between illegal markets for items such as credit cards and legitimate markets for personal data is that illegal data may be considered a "rival good" since the credit card information will be less valuable to criminals if the card has already been used by others – increasing the risk of detection and reducing the potential for illegal use of the card information. This would imply that the market clearing price for a single customer of illegal data is likely closer to the full market value than would be the case with legal data that can be resold multiple times without a reduction in usability for other customers.

Benefits of this approach

The major benefit of this approach is that it provides a price that reflects the market value of a given, specific data entry. Although the market itself is illegal, and thus prices formed here can hardly be compared with prices in legal data markets (see drawbacks), the relative prices of specific data entry as well as the factors affecting their valuation can be instructive. Moreover, black-market values can be better than other valuation techniques at implicitly factoring in externalities.

Drawbacks of this approach

There are a number of drawbacks that limit the usefulness of this approach. First, prices on illegal markets are difficult to collect. Given the illegal nature of the goods and services offered, prices of offers and deals will never be fully transparent and thus are also difficult to measure. As a result, estimations are highly subject to biases due to unrepresentative samples. Furthermore, prices only apply to the context where the data is used again to obtain other benefits illegally. Because criminals must balance the risk of

detection and punishment, the value of the personal data can sometimes be overvalued by such an approach.

Box 5. Data breaches and their economic impacts to firms and consumers

Where personal data are being collected, stored or processed, security incidents can heavily affect privacy as recent high-profile data breaches[8] demonstrate, but they also generate significant costs to firms as well as to users. According to data by the *Privacy Rights Clearinghouse*, the number of incidents identified oscillates around 45 incidents per year, while the total number of records stolen is increasingly determined by large scale data breaches, i.e. data breaches involving more than 10 million records.[9] When combined with the payments arising from pending lawsuits and other measures to reduce the direct and indirect damages caused, the cost per data entry stolen can provide a simplified measurement on the level of risks faced by companies storing personal data in relation to the amount of data entries stored.

The TJX data breach, for instance, involving around 100 million records, forced TJX to set aside USD 118 million to cover costs and potential liabilities in fiscal year 2008 (USD 1.18 per record).[10] This included USD 11 million (9% of total) in security consultancy fees and other attack-related expenses, and a contingency fund of USD 107 million (USD 1.07 per record) to cover liability payments arising from pending lawsuits (see Leyden, 2007). It does however not cover loss in reputation, impact on the brand, and other indirect and opportunity costs. However, TJX announced the impact of the intrusions together with its financial statements for the second quarter (Q2) of 2008. It is interesting to note that the quarterly net income in Q2 2008 was 57% lower than in the same quarter of 2007. This result was probably caused by the higher costs related to the data breach.

Another interesting case is the data breach in Heartland Payment Systems involving around 130 million records in 2009. As a consequence of the breach, Heartland Payment Systems agreed to set up a fund worth USD 105 million to cover liability payments (USD 0.80 per record). USD 41 million (39%) were dedicated to MasterCard customers, USD 60 million (57%) to VISA customers and almost USD 4 million (4%) to American Express customers (see Leyden, 2010). How much Heartland Payment Systems spent on security-related investments as well as the indirect costs is unknown. But the financial statement for fiscal year (FY) 2009 reveals that Heartland Payment Systems had a net loss of more than USD 52 million (compared to a net increase of USD 42 million in FY 2008) even though revenues increased by 7% compared to the previous year. Furthermore, stock prices dropped from USD 18 on 2 January 2009 to USD 15.44 on 16 January and USD 8.54 on 23 January, a drop of almost 50% three weeks (Heartland Payment Systems announced the data breach in 20 January 2009).

A more recent example is the security breach of the Sony's PlayStation Network and Sony Online Entertainment in 2011 which resulted in the exposure of 103 million records and as a consequence to a 23-day closure of the PlayStation Network. According to Sony's executives, this data breach will cost the company at least USD 171 million (USD 1.7 per record). It is interesting to note that this number does not cover liability payments as in the previous cases but rather "expenses of an identity theft prevention programme and promotional packages to win back customers, among other things" (Goodin, 2011); in other words, it covers (parts of) the indirect reputation and opportunity costs. Under the assumption that Sony would also have to set aside a fund worth USD 1 per record to cover liability payments arising from pending lawsuits, additional USD 103 million would have to be provided.[11] This would still not include investments in security assessment and enhancing initiatives (e.g. security consultancy fees).

The cost of data breaches is not limited to the firms suffering from the breach, but also includes the cost consumers have to pay. For example, it has been reported that 10% of Americans have had their identities stolen, and on average, each of those individuals lost around USD 5 000.[12] Similarly, a recent survey of 2 500 Australians found that one in ten had fallen victim to online identity theft in the previous 12 months, but that each had lost an average AUD 1 000 as a result.[13] In the United Kingdom, almost two million people have their identities stolen every year at a cost to the United Kingdom of GBP 2.7 billion. With criminals gaining an average of GBP 1 000 in credit or benefits for each name they steal, the bulk relates to the money criminals obtain using other people's identities with the rest made up of the cost to individuals and companies of preventing and detecting the crime and putting right the damage. In serious cases, it can take people more than 200 hours to resolve the problems caused by identity fraud – the equivalent of a year's annual leave.

Surveys and economic experiments

Another way to put an economic value on personal data is to conduct economic experiments[14] and surveys that extract the price that firms would need to pay individuals to disclose some of their personal information.

In recent years, there have been several experimental studies attempting to quantify individual valuations of personal data in diverse contexts. This area of research is still very not well developed, as the

notion of individual valuation of privacy is very complex and extremely context-dependent, and therefore very difficult to analyse in a laboratory setting. Even though research in this area is still in a preliminary stage, two general messages can be already derived.

- First, people tend to differ with respect to their
    - individual *valuation of personal data* (i.e., amount of money sufficient for them to disclose personal data)
    - individual *valuation of privacy* (e.g., amount of money they are ready to spend to protect their personal data from disclosure)
- Second, these valuations are extremely context dependent, and cannot be measured with an absolute certainty and precision.

The distinction between valuations of personal data and valuations of privacy is of significant practical importance. Concerning the *valuation of privacy*, people do have the possibility to pay to prevent their personal data from being disclosed. For example, Internet users can pay to use software to anonymise their web browsing and to hide their online behaviour but often at the cost of a slower browsing experience. People are also confronted with numerous opportunities to reveal personal information in exchange for some financial benefit, which corresponds to their *valuation of personal data*. For example, some companies offer individuals a bundle of products in exchange for monitoring their Internet behaviour.

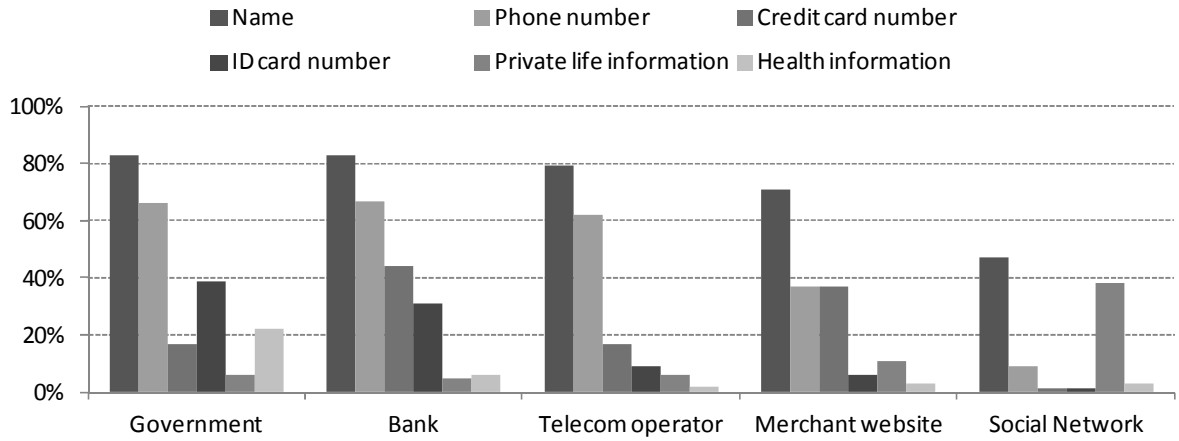
Numerous empirical studies focus on both the *valuation of personal data* and the *valuation of privacy*. Examples include Spiekermann et al. (2002), Chellappa and Sin (2005), Wathieu and Friedman (2005), Huberman et al. (2006), Cvrcek et al. (2006), Hui et al., (2007.) Generally, the fraction of consumers who will reject an offer to obtain money in exchange for reduced privacy is larger than the fraction of consumers who will accept an economically equivalent offer to pay money in exchange for protection of privacy (Acquisti *et al*., 2009). As Hui and Png (2006) noted, the difference between the individual *valuation of personal data* and the *valuation of privacy* could help explain the disparate findings from these empirical studies.

Furthermore, empirical studies point out that both the *valuation of privacy* and the *valuation of personal data* are extremely sensitive to contextual effects. Some of these studies suggest that even ostensibly privacy conscious individuals are likely to share their personally sensitive information with strangers (Spiekermann et al. 2002). An anecdote to illustrate this point is the field experiment presented by Infosec, where 71% of office workers at London's Liverpool Street railway station were willing to reveal their email password for a chocolate bar (Infosec Europe, 2004.) A strict application of the economics principle of "revealed preferences," might lead to a conclusion that people do not care about their privacy nor their personal data (Rubin and Lenard, 2002).

This apparent inconsistency could be explained using the apparatus developed by behavioural economics that takes into account context and psychological individual motives into the economic settings. Using these techniques, several studies such as (Schwarz, 1999; Acquisti *et al.* 2009) demonstrate how significantly the *valuation of privacy* and the *valuation of personal data* can be affected by contextual effects that arguably should play a limited role in decision making. In addition, Acquisti *et al.* (2009) demonstrate the complexity of individual privacy preferences, by exploring the underlying distribution of underlying privacy and personal data valuations. They found that these valuations are not normally or uniformly distributed, but U-shaped, clustered around extremes and focal points.

Surveys also reveal that the impact of the incident depends on the context, in particular the type of information affected by the incident as well as the sector in which the online transaction took place. Some of these surveys also show that users attach less importance to their name or telephone number online, but significantly more importance to credit card numbers, national identity numbers or information about their private lives. The value is particularly high in the case of health-related information (Figure 8).

Figure 8. Importance of personal data according to individuals in France, by attribute type and sector



Source: CDC/IDate survey based on 700 internet users older than 15 years, October 2009

A set of illustrative experiments done by Frog (2011) in the United States, India and China, show there are three tiers of data that individuals are inclined to protect, with the level of protection (and the assigned value) differing by tier: *i*) the top tier includes social security numbers (national identity numbers) and credit card information, which most people value highly (USD 150 – 240 per entry); *ii*) the middle tier contains digital communication history, such as web browsing history, as well as location and health information (around USD 50); and the last third tier of information contains facts about us, including online purchasing history and online advertising click history, to which individuals attach little value (USD 3 – 6).

The same study found that brands and industries are trusted differently. It all depends to whom the data is revealed. Respondents to the Frog (2011) survey declared they put more trust in financial institutions, technology firms, wireless carriers and wireless handset makers. Participants in the study assigned less trust to government agencies and social networks.

Benefits of this approach

Results of surveys and economic experiments have several advantages as a measure of individuals' valuation of personal data. The economic experiments and their procedures are well described in the literature (see Kagel and Roth, 1997), hence this measure is relatively easy to construct and identify. The design of these experiments provides incentives to participating subjects to reveal how they value their data. It means that this measure should be able to capture the pure, unbiased economic value of personal data from an individual perspective.

Furthermore, results usually can be used for comparative studies (across economies and across various types of data). The "sterile" laboratory setting permits to repeat a given experiment on a different sample group and to use the results for comparative studies (across economies and across various types of data).

Drawbacks of this approach

One key drawback of using the results of surveys as a proxy for valuation of personal data is lack of market verification. In fact these measures capture the hypothetical and speculative value of personal data as revealed or indicated by individuals. They capture possible valuations from a subjective, individual perspective, rather than an economic value that was verified by the market forces of supply and demand. These studies have also clearly shown that the values assigned to personal data are highly context dependent. For example, the price at which people are willing to sell their own personal data depends on how they perceive the data will be used and how much they trust the entity receiving the data.

Revealed willingness to pay to protect (insurance)

Another way to put an economic value on someone's personal data is to understand how much the individual would be willing to pay to protect that data. This can be observed in markets offering insurance policies to protect against identity theft. Interestingly, Experian, the data broker mentioned above, sells an identity-theft protection service called ProtectMyID for USD 155 per annum in the United States.[15]

Other firms sell the service of removing people's names from marketing databases. An example is Reputation.com,[16] which allows users to find and remove personal information from websites that they do not, or no longer, want to hold their data. Reputation.com charges roughly USD 140 per year for this service.

Another recent development is the emergence of anti-tracking software and services, which may move the tracking industry towards permission-based tracking and targeting. To date, targeted advertising opt-out mechanisms have been limited, either working rather poorly (e.g., being based on cookies which the aware user will delete and/or not allow to be set), or applying to targeted advertising and not the tracking itself (e.g., offering to stop the targeting, but not the tracking). However, an increasing number of firms offer software and services to limit or thwart tracking. Approaches include relatively simple browser add-ons to block advertising, reveal and manage HTTP and Flash cookies. These add-ons are also available for e-mail clients. Some firms offer preventative solutions, such as SafetyWeb, which offers parental tracking of their children's activities on social networks for USD 100 per year.[17]

Benefits of this approach

A clear advantage of using the revealed willingness to pay to protect the data as a proxy of value of personal data is that it captures the pure economic value of privacy breach from an individual perspective. The value reported by individuals is not affected, by search and transaction costs, nor by market fluctuations and reflects only the pure individual perception of the economic value of potential breach of their privacy.

Drawbacks of this approach

There are two potential sources of bias using the revealed willingness to protect data as a measure for the economic value of personal data.

Similarly to surveys and economic experiments, the revealed willingness to pay for protection lacks market verification. This measure represents the individual's willingness to pay (*valuation*), which is not confronted with the market and not confirmed by market transactions (*value*). For example if a given person reports a relatively low willingness to pay for privacy protection, there might be no insurance company on the market that offers a low enough insurance price.

Secondly, the revealed willingness to pay to protect the privacy captures individually perceived costs of damage caused by data breach rather than value of data themselves. In fact many economic studies highlight the difference between the *valuation of personal data* and the *valuation of privacy* (Acquisti *et al.* 2009; Hui and Png, 2006). Although related with each other, these are two different economic concepts; using measures of *valuation of privacy* to proxy the value of personal data, should therefore be done with great caution.

Finally, there is some debate about whether these services that offer to protect and rectify the theft of personal data are actually effective. This would imply that consumers would be willing to pay less for a service offering questionable results than they may be willing to pay for a more robust service.

Conclusions and next steps

This report marks a first step to understand different methodologies for measuring the monetary value of personal data. It does not look at the economic and social benefits that can emerge from the use of personal data but are not captured in market transactions. It also does not look at other alternative indicators of the personal data market, e.g., the incremental flow of venture capital investments into the market as a result of the potential of data analytics resulting from the use of personal data, or additional investments available for development of new products or services as a result of efficiency gained from the use of data analytics. There are a number of important findings that emerge from this initial study in what could be an ongoing research project.

A need for better data and collaboration

There is a lack of robust data that can be analysed to better understand the economic and social value of personal data. Such an understanding could offer a valuable contribution to the policymaking process. Furthermore, it may be in the joint interest of policy makers and firms throughout the value chain to work together to understand the potential value of personal data.

One key to understanding the potential is better data. Firms in the value chain have a good understanding of the number of records they hold and should have a strategic view of the economic value of these records. There is therefore a valid argument for governments, researchers and firms to work together to better understand the potential benefits and risks of growing amounts of personal data.

Understanding the regional context

More analysis is needed to provide a fuller picture of the economic value of personal data in regional contexts. Currently, most of available statistics that support the presented illustrations of personal data valuations refer to the US market. Therefore, any potential extension of these illustrations onto other markets should take into account regional context, which could include not only legal differences between national privacy regimes, but also socio-economic factors, such as individual awareness, education levels or competitiveness of the marketplace.

The utility of case studies to understand effects

One key theme throughout this report is the context-dependent nature of the valuations of personal data. This entails that macroeconomic effects are much more difficult to analyse because of a lack of harmonised data and measurable impacts over time.

Instead of focusing on macro-economic effects, a partial focus on specific case studies, geared at understanding the effects in a range of contexts, may provide more useful information for policy makers. These case studies could be focused on particular sectors, such as personal data used for advancing health

research. In other cases, they could be tied to specific data sets (e.g., social networking or click-stream data that is used for specific purposes).

Potential non-linear returns means network effects

The monetary, economic and social value of personal data is likely to be governed by non-linear, increasing returns to scale. The value of an individual record, alone, may be very low but the value and usability of the record increases as the number of records to compare it with increases. These network effects have implications for policy because the value of the same record in a large database could be much more efficiently leveraged than the same record in a much smaller data set. This could have implications for competition and for other key policy items such as the portability of data.

Capturing producer and consumer surplus will be difficult

Even if the monetary values of personal data were possible to extract, they would not cover the full economic and social benefit that could be derived from the data. Consumer surplus – the difference between what a user is willing to pay for a good or service and the market price – is not captured in prices in competitive markets. This means that any analysis that omits the consideration of consumer surplus will likely underestimate the true social and economic benefits.

Producer surplus will be captured in the balance sheets of firms using personal data but even this could be difficult to separate from the general operational data of the firm. A focus on specific case studies and closer co-operation with data firms could complement our understanding of the impacts.

Markets where individuals control and sell their own data are evolving and will provide insights

There are new developments on the horizon that could produce new valuations on personal data and shed further light on their market valuations. For example, some firms are now offering "data lockers", which allow users to contribute and edit the data they are willing to share with third parties in exchange for a portion of the proceeds when their data is sold. These data lockers could potentially improve transparency about how data is collected, sold and used. Users may be willing to share even *more* personal data if they feel they have more control over how it is used and received a clear economic or social benefit for sharing. This is a new area and it is unclear if data lockers will emerge with viable business models but this is an area that should be followed.

ENDNOTES

[1] OECD, Terms of Reference for the Review of the OECD Guidelines Governing the Proetction of Privacy and Transborder Flows of Personal Data (2011), available at www.oecd.org/dataoecd/63/29/48975226.pdf.

[2] The final step in the personal data life cycle is disposal, which is a fundamental aspect of the protections related to privacy and information security. It is included as part of this discussion, however, because it does not offer obvious contributions to the value chain.

[3] What They Know: A Glossary, *Wall Street Journal* 31 July, 2010. Available at: http://online.wsj.com/article/SB10001424052748703999304575399492916963232.html (January 2011).

[4] See: http://www.bluekai.com/exchange.php

[5] "Data mining" is defined in many different ways, but is perhaps best understood as encompassing a wide spectrum of data-based activities ranging from "subject-based" searches for information on specified individuals to "pattern-based" searches for unusual or predetermined patterns of activities or relationships. Between these two ends are "relational" searches, which start with an individual but then reach out to determine who communicates or otherwise interacts with whom, and "data matching," which involves combining two or more sets of data looking for matches or discrepancies.

[6] Given the cross-country differences in privacy legislation, and extension of these illustration onto other markets is not straightforward.

[7] Credit card information advertised in cyber crime markets consists of the credit card number, the expiry date, and in some cases even the name on the card (or business name for corporate cards), the billing address, the phone number, the CVV2 number, and the PIN (Symantec, 2010).

[8] A data breach is "a loss, unauthorised access to or disclosure of personal data as a result of a failure of the organisation to effectively safeguard the data" (OECD, 2011b). Where the security breach involve as intellectual property not including personal data the term "unauthorized access" will be used instead.

[9] Malicious hacks still remain the most frequent cause for data breaches in terms of records stolen but not in number of incidents. For example, 63% of all exposed records of incidents recorded between 2005 and 2011 by the *Privacy Rights Clearinghouse* are related to malicious hacks, followed by "lost, discarded or stolen laptop, PDA, smartphone, portable memory device, CD, hard drive, data tape, etc" (*i.e.* lost) with 27%. In terms of number of incidents, malicious hacks rank third (with 19% of all incidents related to this type of events), after loss (31%) and unintended disclosure (20%).

[10] In their 2008 annual report TJX stated that: "[the] actual losses arising from the Computer Intrusion could exceed [their] reserve for [their] estimated probable losses, and [their] reputation and business could be materially harmed as a result of any future data breach".

[11] It has been suggested that this breach may have cost card issuers alone between USD 230 million and USD 385 million for card replacement if all customers affected by the breach decided to replace their credit

cards. This estimation is based on the cost for replacement which is expected to be between USD 3 and 5 per card, and the 77 million exposed accounts (see Reuters, 2011).

12 http://mashable.com/2011/01/29/identity-theft-infographic/.

13 AAP (2010) 'ID theft costs $1.3bn,' *The Australian* 6 July 2010. Available http://www.theaustralian.com.au/australian-it/id-theft-costs-13bn-survey/story-e6frgakx-1225888567247 (September 2010).

14 Economic experiments are designed to study people's behaviour in some economically-interesting scenarios using monetary incentives. In doing so the experiments mimic some features of real-world situations in a stylized, laboratory setting.

15 www.experian.com/consumer-products/identity-theft-protection.html.

16 www.reputation.com/.

17 http://www.safetyweb.com/.

REFERENCES

Acquisti, A., L. John, and G. Loewenstein (2009) What is Privacy Worth?, mimeo, available at: http://pages.stern.nyu.edu/~bakos/wise/papers/wise2009-6a1_paper.pdf, accessed on 18 May 2012

Anderson, C. (2009), "Free: The future of a radical price", Hyperion. Available http://www.longtail.com/the_long_tail/2009/07/free-for-free-first-ebook-and-audiobook-versions-released.html (November 2010).

Angwin, J. (2010), "The Web's New Gold Mine: Your Secrets", *Wall Street Journal* 30 July, 2010. Available http://online.wsj.com/article/SB10001424052748703940904575395073512989404.html

Angwin, J. and McGinty, T. (2010) "Sites Feed Personal Details To New Tracking Industry", The Wall street Journal, 30 July 2012, Available http://online.wsj.com/article/SB10001424052748703977004575393173432219064.html

Angwin, J. and Steel, E. (2011) "Web's hot new commodity: Privacy", *The Wall Street Journal* 28 February 2011. Available http://online.wsj.com/article/SB10001424052748703529004576160764037920274.html?mod=WSJ_Tech_MIDDLETopNews.

Appel, J.M. (2008), "Why shared medical database is wrong prescription", Orlando Sentinel, December 30, 2008 http://en.wikipedia.org/wiki/Electronic_health_record

Beales, H. (2010), "The Value of Behavioral Targeting", Network Advertising Initiative, Available http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf

Berman, S.J. (2011), "Not for Free: Revenue strategies for a new world", Harvard Business School Press.

Berman, S.J., Battino, B. and Feldman, K. (2010), Beyond content: Capitalizing on the new revenue opportunities, IBM Institute for Business Value, Available *http://ibm.com/iibv* .

Boyd, D. (2010), "Privacy, Publicity, and Visibility", Microsoft Tech Fest. Redmond, March 4. Available www.danah.org/papers/talks/2010/TechFest2010.html (October 2010).

Burridge, N. (2010), "Annual cost of identity theft is £2.7bn", *The Independent*, 18 October 2010. Available www.independent.co.uk/news/uk/crime/annual-cost-of-identity-theft-is-16327bn-2109431.html (April 2011).

Cate, F.H. (2008), "Government Data Mining: The Need for a Legal Framework", in *Harvard Civil Rights-Civil Liberties Law Review*, Vol 43. Available http://harvardcrcl.org/ (September 2011).

Chellapa, R. and R.G. Sin (2005), "Personalization Versus Privacy: An Empirical Examination of the Online Consumers' Dilemma," *Information Technology and Management*, 6(2-3), 181-202.

Cvrcek, D., M. Kumpost, V. Matyas, and G. Danezis, 2006. "A Study On The Value Of Location Privacy," Proceedings of Workshop on Privacy in the Electronic Society (WPES '06), 109-118.

CyberSource (2011), 2011 Online Fraud Report, CyberSource. Available www.pymnts.com/cybersource-2011-online-fraud-report-12th-annual-edition-online-payment-fraud-trends-merchant-practices-and-benchmarks (April 2011).

FTC (2009), "Self-Regulatory Principles for Online Behavioral Advertising", FTC, Washington DC. Available www.ftc.gov/opa/2009/02/behavad.shtm (September 2010).

Frog (2011), "The Value of Personal Data: A Global Perspective", Presentation at Mobile World Congress, February 2011.

Goodin, D. (2011), "PlayStation Network breach will cost Sony $171m", The Register, 24 May 2011, Available http://www.theregister.co.uk/2011/05/24/sony_playstation_breach_costs/

Hui, K.-L., H-H. Teo, S.-Y. Lee (2007). "The Value of Privacy Assurance: An Exploratory Field Experiment," *MIS Quarterly*, 31(1), 19-33.

Hui, K-L. and Png, I.P.L. (2006) The Economics of Privacy, in Hendershott, T. (ed.) Handbooks of Information Systems, Volume 1. Elsevier.

Infosoc (2004), "Passwords for a Chocolate Bar" survey; summary available at: www.net-security.org/secworld.php?id=2075

Kagel, J. H. and A. E. Roth (1997), "The Handbook of Experimental Economics", Princeton University Press

Leyden, J. (2010), "Heartland coughs $41m to settle MasterCard claims", The Register, May 20 2010 Available http://www.theregister.co.uk/2010/05/20/heartland_mastercard_settlement/

Narayanan A. and V. Shmatikov (2010) Privacy and security Myths and fallacies of "Personally identifiable information," Communications of the ACM ,vol. 53 (6)

Neate, R. and Mason, R. (2009), "Networking site cashes in on friends", *The Telegraph*, 31 January 2009. Available www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/4413483/Networking-site-cashes-in-on-friends.html (April 2011).

OECD (2010), The role of Internet Intermediaries in Advancing Public Policy Objectives, Workshop, Paris June 2010, OECD DSTI/ICCP(2010)13.

Ohm, P. (2010), "Broken promises of privacy: Responding to the surprising failure of anonymization". *UCLA Law Review*, 57

Panda Security (2010), "Panda Security Report. The Cyber-Crime Black Market: Uncovered", Available at: http://press.pandasecurity.com/wp-content/uploads/2011/01/The-Cyber-Crime-Black-Market.pdf

PWC (2010), IAB Internet Advertising Revenue Report: 2010 First Half-Year Results, the Interactive Advertising Bureau, October 2010. Available at www.iab.net/insights_research/947883/adrevenuereport (October 2010).

Schneier (2010), SecuritySchneier on Security, A blog covering security and security technology. Available at: http://www.schneier.com/blog/archives/2009/11/a_taxonomy_of_s.html

Schwarz, N. (1999), "Self-reports: How the questions shape the answers," *American Psychologist*, 54(2).

Sinclair, L. (2010) "Facebook targets $2bn as it overtakes Google", *The Australian*, 22 March 2010. Available at /www.theaustralian.com.au/business/media/facebook-targets-2bn-as-it-overtakes-google/story-e6frg996-1225843478239 (September 2010).

Spiekerman, S., Grossklags, J. and Berendt, B. (2002), "E-privacy in 2nd generation E-Commerce", ACM Conference on Electronic Commerce (EC'01), Hrsg. ACM New York, 38-47. Tampa, Florida: ACM Press.

Steel, E. and Angwin, J.(2010), The Web's Cutting Edge, Anonymity in Name Only, *Wall Street Journal* 4 August, 2010. Available at: http://online.wsj.com/article/SB10001424052748703294904575385532109190198.html (November 2010).

Symantec (2010), "The Silent Epidemic: Cybercrime Strikes More Than Two-Thirds of Internet Users", Press Release, 8 September. Available at: www.symantec.com/about/news/release/article.jsp?prid=20100908_01

Symantec (2011), "Symantec Internet Security Threat Report: Trends for 2010", Volume 16, April.

Thurm, S. and Yukari, I.K. (2010), "Your Apps Are Watching You", *Wall Street Journal* 17 December 2010. Available at *http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html?mod=what_they_know* (January 2010).

Tucker, C. (2010) The Economic Value of Online Customer Data, OECD, Paris. Available at *www.oecd.org/document/22/0,3746,en_2649_34255_46565782_1_1_1_1,00.html* (January 2011).

Tynan, D. (2007) As Applications Blossom, Facebook Is Open for Business, Wired Magazine, July 2007. Available *www.wired.com/techbiz/startups/news/2007/07/facebook_platform* (November 2011).

Wathieu, L. and A. Friedman, 2005. "An Empirical Approach to Understanding Privacy Valuation," Proceedings of the Fourth Workshop on the Economics of Information Security (WEIS '05).

WEF (2011) Personal Data: The Emergence of a New Asset Class, World Economic Forum. Available *www.weforum.org/issues/rethinking-personal-data* (May 2011).