Earth, Life & Social Sciences
Van Mourik Broekmanweg 6
2628 XE Delft
P.O. Box 49
2600 AA Delft
The Netherlands

www.tno.nl

T +31 88 866 30 00
F +31 88 866 30 10

TNO report

TNO 2014 R11390 | Final report

Personal Data Markets

Date	2 November 2014
Author(s)	Arnold Roosendaal, Marc van Lieshout, Anne Fleur van Veenstra
Copy no	-
No. of copies	-
Number of pages	57 (incl. appendices)
Number of appendices	-
Sponsor	TNO
Project name	Personal Data Markets
Project number	060.03591

Contents

1 Introduction

This is the final report of the Personal Data Markets (PDM) research project carried out by TNO. The PDM research project was an exploration of PDM and the relation with personal data management. Besides, the aim was to get an overview of what a PDM ecosystem looks like and how it works in practice.

The intensity of personal data processing has grown exponentially over the past years. Developments in telecommunication technologies and electronic service delivery facilitate massive processing of data, while at the same time providing opportunities for new business models. The economic value of personal data processing has been widely recognized. In the field of regulation, this brings specific challenges, which can also be seen in the big legal data protection reform that is currently taking place at a European level.

Thus far, research has mainly focused on the legal requirements for legitimate personal data processing. Also economic chances have been sketched quite well. The availability of personal data for new services in combination with a proper protection of the interests of the individual consumer, however, remained relatively out of scope. There have been some initiatives where consumers were provided with the opportunity to decide for themselves whether they allow the usage of their personal data, by whom, and for what purpose, sometimes even with a financial reward. Nevertheless, a broad application of services is not yet taking place. With an eye on the added economic value and adequate protection of the interests of consumers, it is necessary to gain more insight in which actors are needed and which requirements apply for successfully implementing personal data market ecosystems in practice.

The results of the research are presented in this report. First, in chapter 2, an overview of current literature on personal data markets, the estimated value of personal data and presents some findings on how individuals value their personal data is presented. Then, an approach towards a PDM ecosystem and the main actors involved and what such an ecosystem looks like in a simple form is presented (chapter 3). In chapter 4, a number of applications of personal data in relation to different types of data categories that can be discerned are outlined. Subsequently, a number of case studies on PDMs that can be found in practice are described in chapter 5. The selection is necessarily limited and incomplete, but presents an interesting overview of different modalities of PDMs. It starts with a PDM constructed by an individual data subject, then discusses the emergence of PDMs in healthcare on the basis of a number of institutes that deal with large patient data sets, and finally discusses some PDMs that emerged in the financial sector recently. The findings of the previous chapters are enriched in chapter 6 by adding a view on privacy features for PDM. Finally, in chapter 7 the main findings and conclusions of the study are presented.

2 An overview of personal data markets

2.1 Introduction

Personal data are an important asset in today's data-driven society. With regard to personal data, often a privacy paradox is observed with individuals stating that they find privacy to be important, but not acting accordingly. The sharing of personal information on social networking sites, blogs and websites seems to support this view. Several arguments can be presented for this apparent contradiction. For instance, users do not have a choice but to accept that they hand over their personal data in exchange for the service to be delivered. Or, users make a distinction between what they consider privacy relevant and what not which deviates from the formal distinction between 'ordinary' personal data and sensitive data. Or, they seem to share quite some personal data, but they use informal strategies to shield their privacy (for instance by supplying incorrect personal information). Another perspective takes the awareness of economic value related to this data as a starting point, which has triggered the idea that individuals should be able to maintain control over their data and even make their own profit from their personal data (OECD, 2013). This user-centric control approach is usually described by the umbrella term 'personal data management' (Hildebrandt et al., 2013). The profit part concerns the monetization of personal data. In order to make money by selling or licensing the use of personal data, some form of a market is needed where individuals can offer their data and where companies willing to use the data can place their demand. The precise organization of the exchange will have its influence on issues such as privacy. We approach this exchange as taking place within an ecosystem of actors and relations between actors in a sustainable manner. What such an ecosystem may look like and what roles are needed and which requirements should be met is the core question of this report.

Literature in the field of behavioural economics shows that there is a very limited willingness to pay for the protection of privacy (Acquisti, 2009). Neither do people value their personal information as being very worthwhile when related to ordinary personal information (Spiekermann, 2012). The relationship between data and privacy is not linear either. People tend to value what they have above what they may gain (endowment effect, Thaler, 1980). Other research shows people do not esteem their personal data as very worthwhile as long as this concerns data they share through social websites. The question that arises is whether individuals are searching for the real cash that may lay hidden in their personal data, or rather for a trustworthy system for exerting control and having a certain level of transparency. To complicate matters, determining the monetary value of personal data appears to be rather difficult. Various ways of quantifying data can be used, varying from the (stock) value of companies which process personal data as their core business, such as social networking sites, in relation to the number of members they have, to market prices for data, illegal markets, and surveys (OECD, 2013). Different methods lead to different outcomes. The monetary value of personal data appears to have a high level of context-dependency, which makes it difficult to capture macroeconomic effects (OECD, 2013). As a result, the pricing schemes that may help in creating personal data markets may not be that straightforward to develop. Moreover, a critical approach of the monetization of personal data needs to

question whether monetizing personal data will be able to weaken the fundamental rights people have on the protection of their privacy. Our intuitive stance towards this is that it is problematic to 'trade' fundamental rights away by offering money for personal data. Before tackling this more fundamental issue we will first discuss what precisely we mean with personal data markets and what models can be used to chart the value of personal data within these markets.

2.2 Literature review: the market for personal data

A number of recent studies have been published in which personal data markets have been investigated (WEF, 2013, WEF, 2010, BCG, 2012, Deighton and Johnson, 2013) The studies differ in what they precisely investigate, ranging from the overall impact of changes in the use of personal data (WEF 2010, 2013) to a study on the size and the features of the European markets on personal data (BCG 2012) and a US-based investigation of the market in which Individual-Level Consumer Data play a pivotal role (Deighton and Johnson, 2013).[1] The WEF 2010 study highlighted the relevance of personal data as an economic asset that could be perceived as the new 'oil'. The metaphor of personal data as oil is an interesting one. It covers both the use of oil as a product in itself and as being a substance that is basic to a large number of economic activities, and that itself should be considered as 'raw' material or 'semi-finished product'. The WEF study was one of the first in trying to come to terms to the phenomenon of 'big data' developments, and identified a number of interesting features of these developments. It introduced an - arguably contestable – distinction between types of personal data that enables a kind of classification of the data processing and collecting processes (based on either voluntary, observed or inferred data, see below).

The Boston Consulting Group (BCG, 2012) presented a study in which it tried to estimate the economic value of personal data markets in Europe, introducing a new inventory of relevant economic sectors. The study arrives at an expected value of personal data markets of 8% of European GDP in 2020. It bases its forecast on a composite average of growth of some prominent economic sectors at present. The three most relevant ones are online communication and entertainment (CAGR 22%), ecommerce (CAGR 15%) and web-communities (100%). Given the present market size a presumed average CAGR of 22% over the next years would yield a market of €330 Billion in 2020,while consumer benefits would even be bigger (€670 Billion) due to reduced prices, time savings (because of self-service transactions) and the valuation of free online services.

The US-based study on Data Driven Market Economies, published by the Data Driven Marketing Institute, on the role of ILCD in providing marketing services shows that producers of goods and services spent about $156 Billion and employed 676.000 people for marketing services on the basis of Individual Level Consumer Data (ILCD). With a total US marketing and advertising market of $298 Billion the contribution of ILCD based marketing is roughly 50%. The largest contribution to the economic value of DDME (slightly over 70%) is related to the direct (50%) and

[1] Many other scientific publications have considered the economic value of personal data since many years. See for instance Acquisti and Grossklags, 2007 and Spiekermann, 2012. In recent years the search for the value of personal data in a market that is determined more and more by large datasets (the big data revolution) has yielded to increased attention.

indirect (21%) exchange of ILCDs between firms, with only some 29% related to the collection of use of ILCD within a single firm (Deighton and Johnson, 2013) (p. 15-16). The role of personal data in marketing activities is manifold, and ranges from personalised targeting, to measuring benefits of marketing activities and lowering the entry costs for small firms (for which mass-advertising is too costly). The study indicates as well the relevance of DDME for stimulating technology development and realising start-up entrepreneurship.

The relevance of personal data has a public element as well. Personal data are a prime asset for public services. Having access to (reliable) personal data may improve efficiency of public services. But the relevance of collecting, aggregating, analysing and using personal data is more than just can be expressed in monetary terms. An example is provided by Kaiser Permanente, which has collected a database with over 3 million patient records. It offers patients fast access to their medical files, allowing them for instance to schedule an appointment and to receive text messages for prescription refills, leading to cost savings for Kaiser Permanente of many hundreds of thousands of dollars (Katibloo, 2011). Meanwhile, the same data can be used to investigate correlations between incidences of diseases and use of medicines. To give one example, Kaiser Permanente was able to identify a correlation between use of anti-depressants by pregnant women and the incidence of a form of autism by newborns (WEF, 2013, p.8). These results enable adapting medical practices in anti-depressant prescriptions for pregnant women.

Massive personal data collection can thus serve multiple purposes. The area identified in the US study relating to advertisement networks in online environments can be seen as a growing and interesting part of personal data markets. The market is growing with the growth of personal data that people leave, knowingly or unknowingly, when using one of several digital platforms they have at their disposal. An estimate presented in the BGC study mentions a growth figure of 45% per year through 2015 to a volume of 7 zettabytes, being the equivalent of more than 1.000 gigabytes of data for each person on earth (BCG, 2012). The advertisement market is an interesting domain since it has already matured to some respect, being endemic to the growth of a large number of 'Freemium' services in exchange for more detailed information on personal belongings.

A market of personal data brokers has developed who focus on the delivery of collected and enriched personal data to customers interested in this information for their own interests. Large service providers such as Google, AOL and Yahoo have in the recent past taken over information brokers to secure their own position on this market. Google houses a number of brokers such as AdMob and recently incorporated Double Click. Apple has its own ad-broker with iAd. The role of these ad-brokers is growing. The ad-broker MobClix for instance, matches 25 advertisement networks with 15.000 different apps that are seeking advertisers. The advertisement market on mobile platforms (tables, mobile phones) has matured and has led to the rise of organisations such as BlueKai that offers a data exchange platform that captures more than 30.000 attributes over 300 Million users. Its activity surpasses 75 Million auctions a day (an auction being the means with which advertisement space is offered to potential advertisers; ad networks are the intermediaries between advertisers and those offering advertisement space) (OECD, 2013).

2.3 The value of personal data

The digitisation of communication and information has given rise to an abundance of data-sharing practices. People share details about their whereabouts, their moods, their activities, through a multitude of platforms. They leave traces that go unnoticed for themselves, such as their geo-location when carrying a mobile phone, or their click behaviour. The value of this information is well understood by marketers who try to collect as much data about personal conducts and preferences as possible, allowing them to learn about purchasing habits and strategies, to create the best profiles possible, and to make the best suited offers to their customers. As indicated above, it is not only the commercial value of personal data that is of interest but the public value as well. Data which are hardly relevant from a commercial point of view may serve specific public purposes that make them worthwhile. An example is the delivery of medical information by patients with rare diseases.[2] The network of patients with rare diseases started as a social interaction between these patients but resulted in extremely interesting data exchange for medical practitioners (and pharmaceutical agencies), When trying to grasp the meaning of the value of data it is prerequisite to distinguish between the kinds of values that data may embody. In the following we will start from the economic and monetary value personal data can have. Having presented an overview of how to assess this monetary value we will also pay some attention to the assessment of public value of personal data.

2.3.1 The monetary value of personal data – a market perspective

To measure the economic, or monetary, value of personal data, two main perspectives can be used. The first is by assessing the monetary value of the firm that collects, aggregates, processes, stores and/or disseminates the personal data. Various approaches are possible for this assessment (OECD, 2013). The second perspective determines which monetary value persons attach to their data. This can be assessed in various manners as well.

To start with the monetary value of personal values from a firm's perspective, a recent OECD study distinguishes between three perspectives: one can look at the stock value of a firm, at the revenues of a firm or at the price of data records on the market. Alternatively, one can also look at the costs of a data breach and at the price of personal data on an illegal market. All of these approaches show some features of the value of personal data but all have specific drawbacks as well.

A general feature of data is that it can be sold over and over again without loss of its intrinsic value (such as showing the birth date of a person). The copy is just as good as the original, enabling multiple offers without loss of price or value. This presupposes that data do not get their value because of being exclusive but merely because of them being available. This introduces a disruptive element in business models that are based on exclusivity of what is sold. Another relevant feature is the opportunity to use personal data for the creation of profiles in which specific filter categories determine to what specific profile someone belongs. Profiles can be created bottom up (using the available data to create meaningful subsets of data) or top down (using pre-configured profiles to check in what group specific people

[2] See www.patientslikeme.org

would belong). Both forms of profiles add to the monetary attractiveness of personal data, since the grouping of data add to the original value of the data.

The *stock value* of a firm is a measure of trust in the firm's capacity to produce valuable revenues. It expresses the expectation shareholders have in the growth potential of the firm. For firms trading in personal data as their primary source of revenues, the stock value may be used as a proxy for the value shareholders attach to the data collected and the processes that turn the data into profitable products. Stock values may however fluctuate depending on contextual factors that do not bear direct relationship with the primary process of the firm. Trust in firm's behaviour may rely on such issues as a CEO who leaves the firm due to private circumstances. This in itself has no direct link with the value of the data the firm collects, aggregates, processes and distributes. Fluctuations of stock prices can induce further fluctuations, as was shown by the introduction of Facebook to the stock market. Only in relatively stable markets one might expect a relatively stable relation between the value of a firm's shares and the revenues it realises on the basis of its business activities.

The *revenues of a firm* may serve as a better proxy since it indicates real cash flows on the market, due to the firm's ability to sell products to customers. It enables cross-comparison between firms acting on a similar market, since one would expect these firms to encounter similar problems in selling their products. The revenues per record may be an indication of the ability of a firm to overcome the complexity of the market, yielding higher revenues against lower costs. Revenues should be compared to the total number of data records a firm owns in order to yield a comparative indicator (revenues per data record in a specific period of time). A drawback of this method is that external factors may influence the prices third parties are willing to pay for specific data on the market, and that there may be a dependency on the total number of records a firm possesses (synergistic effects due to the fact that a firm is able to offer a larger sample of personal data records).

When a firm not only processes and sells personal data but combines this with other activities, the revenues per record may be blurred through these other revenues as well. Net revenues per data record (profits minus costs) deliver a better measure than pure revenues (not including costs). Costs can vary considerably between firms and between markets. This may influence the value of personal data as well.

An example that shows the variance between the indicators above is provided by Experian (OECD, 2013). Experian is a data broker. Over 2011, Experian reported total revenues of USD 4,2 billion realised over 600 million individual and 60 million business data records. Its stock value circled around USD 10 to 12 billion. Market capitalisation thus is about USD 19 per record, and annual revenues were about USD 6 per record. Profits were roughly USD 1 per record. By means of comparison, the stock market prices of Facebook have seen huge fluctuations since its introduction (from an initial USD 38 to a low USD 20 low per stock two months later to a value of USD 55 in December 2013[3]. Market capitalisation of Facebook developed from some USD 90 billion at the start of its presence at the stock market

[3] http://en.wikipedia.org/wiki/Initial_public_offering_of_Facebook (accessed January 7, 2014)

up to USD 140 billion in December 2013.[4] Over the past four quarters (Q4-2012 up to Q3-2013), Facebook earned a total of USD 6.9 Billion (with USD 2,0 Billion in 2013 Q3) and had a profit of USD 1,04 (with USD 0,43 Billion in 2013 Q3).[5] Over this year (Q4 2012 – Q3 2013), Facebook has a market capitalisation of USD 116 per subscriber, revenues of USD 5,75 per subscriber and a profit of USD 0,87 per subscriber. Though not all revenues are due to the selling of ads, a large part is. The market capitalisation of Facebook thus is much and much larger than the market capitalisation of Experian while other indicators are in the same range (with the note that positive revenues and profits for Facebook only started at Q4 2012).

Table 1: Comparison between Experian and Facebook on indicators of value of personal data, captured by these firms (OECD, 2013; TNO 2014)

	Experian 660 M users		**Facebook 1,1 B users**	
	Total value	Per record	Total value	Per record
Market Capitalisation	$10-12 billion	$19 (2011)	$90-140 billion	$110 (Q4-2012; Q3-2013)
Revenues	$4 billion	$6	$6.9 billion	$6,25
Profit	$660 million	$1	$1,04 billion	$0,92

The price of personal data as these are sold at the market place offers another indicator. In this situation it depends on the value potential purchasers of personal data attach to these data, which in turn will depend on the profitability they expect to realise. In a 2013 web-based article the Financial Times offers an interactive sheet that enables calculating market prices for specific sorts of data.[6] It distinguishes between demographic data, family and health data, property, sport and leisure activities and consumer data. Demographic data as age, gender, ethnicity, zip-code and education level are worth USD 0,005 per piece. Job information is worth USD 0,1 if being an entrepreneur up till USD 0,72, if being a health professional, pilot or non-profit worker. Over the five data categories, a total of 24 data entries can be discerned, each worth a specific (usually very modest) price. In a response to the FT article, it was argued that specific personal items have a much higher value, again depending on the kind of information searched for. Having information on credit history, criminal records, bankruptcies, convictions etc. of persons can cost up till USD 30-40 per record.[7] Firms specialise in inquiries for this kind of background information. Since this is on particular persons and not on groups or

[4] http://en.wikipedia.org/wiki/Initial_public_offering_of_Facebook (accessed January 7, 2014)
[5] http://techcrunch.com/2013/10/30/facebooks-q3-13-beats-with-2-02b-revenue-0-25-eps-with-49-of-ad-revenue-now-mobile/ (accessed January 7, 2014)
[6] http://www.ft.com/intl/cms/s/2/927ca86e-d29b-11e2-88ed-00144feab7de.html (accessed January 7, 2014). BTW: in order to access these pages one has to register oneself, thus adding to the value FT derives from its subscribers!
[7] https://ioptconsulting.com/ft-on-how-much-is-your-personal-data-worth/, referring to http://backgroundreport360.com/ (both accessed January 7, 2014).

cohorts, prices are much higher, this presenting a different kind of market. This still relates to legally available information. Information that is available on black market prices shows that data on credit card numbers, personal health records and the like may cost USD 1-30 per record, depending on the sensitivity of the data but also on the occurrence of data breaches (which provide new data on the market but also may lead to a saturated market) (OECD, 2013).

Data breaches may themselves offer another inroad to measuring the value of personal data. A data breach as occurred to the Sony Playstation Network between April 17 and April 19 2011 led to the theft of personal data of 77 million subscribers to the Sony Playstation Network. It led Sony to stop its services for 24 days. Together with the costs of recovering from the hack and the fines to be paid, the data breach cost Sony USD 171 million, this being the directly attributable costs. Per subscriber this leads to a value on the side of Sony of USD 2,20. The indirect costs (loss of subscribers, negative brand image which may lead to a decline of purchases of other equipment as well, impact on stock market prices) have been estimated at USD 1,25 billion, being USD 16 per subscriber.[8] Stock market prices showed a dip when Sony entered the stock market again of some 6%, but it is hard to decide whether this is due to the data breach or to the overall fall of stock market prices that Sony experiences in the period February 2011 – November 2012 (steadily falling down from USD 37 up to USD 9 over this period).[9]

From the above, it can be concluded that there is some value in having prices per data record. However, there is very much dependence on the typical situation, since, usually, the value is in the mass. A simple translation to price per record can, thus not always be made. Moreover, it is difficult to say whether the commercial price of a data record is of relevance for a firm to decide whether it should enter a market or not, and what it takes to realise a break-even point. It seems that firms with a large user database have an advantage over smaller firms. However, when the market for personal data matures, this advantage may disappear, because the individual value per record may become of more relevance, depending on the type of service provided by a firm.

2.3.2 The monetary value of personal data – a 'personal' perspective

A second approach to measuring the monetary value of personal data is by assessing the value individuals attach to their personal data. This is an approach that may bear relevance for studying the presumed changing role of privacy for individuals. One could argue that individuals will value personal data that explore more sensitive issues about personal traits, habits, preferences, attitudes, and activities higher than data that they do not consider to reveal anything interesting. By measuring at which value individuals would be willing to release this information for commercial or other purposes one could investigate at which price people are willing to lift their privacy for specific data categories.. This is however not a straightforward approach. For one, the notion of sensitivity of data is a relative one. Data that are not considered to be sensitive in one situation may be very sensitive in another situation (context dependability). Date of birth, for instance, does not

[8] Juro Osawa, May 9 2011. 'As Sony counts hacking costs, analysts see billion-dollar repair bill.' *Wall Street Journal*. http://online.wsj.com/news/articles/SB10001424052748703859304576307664174667924 (accessed January 7, 2014).
[9] http://quotes.wsj.com/SNE/interactive-chart (accessed January 7, 2014).

appear to be an issue except for when one needs to prove to be eighteen or older and except when one indeed has grown much older. Home address is another example. In ordinary situations one may not care about who knows where one lives but in specific cases (such as women seeking shelter from violence in home situations) the home address becomes a very sensitive datum. For another reason, and as indicated in the introduction, we question the assumption that individuals will assess their personal data purely as an economic or monetary asset, as a tradable good under all circumstances. They might value other aspects related to their data as more relevant, such as the ability to exercise control over their data. Finally, by presupposing that people are willing to lift part of their privacy by revealing personal data, one implicitly presumes that privacy is just the protection of personal data. Privacy is however a broader concept (Solove, 2008; Finn et al, 2013). Privacy not only deals with the person but has a social relevance as well (Regan, 1995). Being able to exploit one's unique opportunities without interference of others is in this context as much part of one's privacy as "being able to determine for oneself in what situation what information is shared with others" as an often quoted description of privacy goes (Westin, 1967). The protection of personal data, on the contrary, is much more a defensive concepts that refers to preventing that specific data will become available for specific organisations and/or for specific uses.[10]

The role of personal data in today's society is undisputed. In a Eurobarometer Survey, stemming from 2011, 74% of respondents indicated that they accept personal data need to be disclosed when participating to today's society (TNS-NIPO, 2011). The same number of people consider financial information, medical information and identity card numbers as personal information. Differences between respondents can be found in terms of educational level and geographical origins, with higher educated people and people living in Western and Northern European countries more represented in the identification of specific data categories as personal data (TNS-NIPO, 2011).

Statistical analysis does not show motives or preferences towards privacy. In understanding what people value in privacy the traditional economic models, based on a rational actor, have been supplemented with models that look at behavioural features. These models study the impact of attitudes and preferences on economic choices people tend to make. They start from an individual who is subjected to a complex mix of influences of an economic and a non-economic character. Within behavioural economics several of these influences have been studied (Acquisti, 2009; Rabin, 2013; Spiekermann, 2012). Information asymmetry influences behaviour: in many situations the access to relevant information is different for different actors.. People tend to value benefits differently from losses (hyperbolic discounting, (Rabin, 2013). People tend to value what they own higher than what they not own (endowment effect, Thaler, 1980). People tend to overvalue immediate rewards and undervalue long term rewards (instant gratification). The absence of real choices may impact upon how people will behave. When one only can chose between accepting specific conditions and getting access to a service or rejecting the conditions and thus having no access to that service, one may be

[10] On top of this difference, Gutwirth and Gehlert demonstrate that the protection of personal data is mostly embedded in procedural rules that do not impose any specific meanings of the concept privacy while the protection of privacy, as it reaches the European Court of Justice in order to decide whether privacy is infringed yes or no, is always embedded in a substantive argumentation, paying due respect to the precise circumstances of the infringement. (see Gutwirth et al, 2011).

tempted to accept unfavourable or unclear conditions. This practice is well-known in the internet-economy. Many services are offered as a 'take it or leave it' option. For many youngsters it is absolutely prerequisite to have a subscription to Facebook, if one wants to keep in close contact with one's friends, and thus one has to accept the conditions Facebook poses, whether one likes this or not.

Empirical research that tries to identify the relevant parameters of behaviour in order to understand how people assess the value of their personal data, is relatively scarce. A recent study by ENISA identities four papers dealing with an empirical field- or lab-related study of privacy behaviour (ENISA, 2012). On top of the studies reviewed, ENISA performed itself a case study in which it studied whether people are willing to pay for additional data protection (ENISA, 2012). When buying a ticket for the cinema, participants could choose between a number of – varying – offers. Minimum set of data asked was name, e-mail address and date of birth. Variations existed in the usage of the data (using e-mail address for advertisement options) and request for additional information (phone number). On top of this in some experiments the price was kept the same for different options while in other options the price was different between the privacy-friendly and the privacy-unfriendly firm. The experiment was conducted as a lab experiment (with 443 participants), in which different options were offered in sequence, and as a hybrid field experiment (with more than 2.300 participants). The study showed that the privacy-unfriendly option was chosen by the majority of the participants when the ticket could be bought by a price reduction of 50 cent compared to the privacy-friendly offer. A minor part of 13% chose to pay the additional 50 cents. Without price difference, the majority of participants chose the privacy-friendly option. The experiment also showed that participants when buying two tickets consecutively, remained to a large extent (142 of 152 participants) loyal to their first choice, even when the second choice meant they could pay less for their ticket (ENISA, 2012).

In a 2007 paper and Acquisti and Grossklags present the results of their research into the willingness of people to pay for data protection vis-à-vis the willingness of people to sell data against a similar price. They are interested in studying whether the gap between 'Willingness-to-Pay' and 'Willingness-to-Accept' as identified in several studies, exists as well when dealing with privacy and data protection issues.[11] The Willingness-to-Pay for the protection of personal data was relabelled as Willingness-to-Protect. This was juxtaposed against the Willingness-to-Accept a financial reward in return for release of specific personal data. The experiment of Acquisti and Grossklags performed is much more modest than the ENISA studies. The findings are thus illustrative at most. They had participants (47, mostly students) first filling in a quiz and answering some additional and sensitive issues, one on the number of sex partners the participants had had (this being the most sensitive bit of information participants exposed). Having the information, participants could either protect information (at a specific cost) or sell it. Situations were such that they enable comparisons between the different situations (also in terms of monetary situations). The experiments validated the gap found between WtP and WtA. More research is however needed to master all possible conditions that may influence the decision making process.

[11] This gap is identified in many studies. People tend to value what they own above what they do not own and are thus willing to pay a higher price for keeping what they have than for achieving the same when not having it. See Acquisti and Grossklags and Acquisti, 2007.

A final study worth mentioning is a study performed by Sarah Spiekermann in which she investigated the willingness of participants to pay for their own data which they had left at Facebook before (Spiekermann, 2012). The – hypothetical – situation Spiekermann sketches is the announcement by Mark Zuckerberg that he pulls the plug out of Facebook. He offered all who had information on Facebook to either buy their information back or have it destroyed. In another option (sketched by Spiekermann) participants had the option to have their data sold to a third party or to buy their data in order to prevent the selling. In a third option people had the option to have their data sold to a third party but to share in the revenues. The experiment was performed with over 1.500 Facebook participants. The results of the experiment showed that the Willingness to Pay was lowest in the first option (money one was willing to pay in order to have data saved): ~€16,- (median). In the second option Willingness to Pay was higher: €54,- (median) for preventing the data were sold to a third party. In the option of sharing in the revenues, people assessed the value of their data considerably higher: €507,- (median). Spiekermann analysed the propensity of people towards their data. She concludes that people build up a psychology of ownership that becomes triggered the very moment they realise the value hidden in their data. Psychology of ownership is more relevant than privacy concerns in explaining attitudes of people vis-à-vis their personal data.

2.3.3 Conclusions

Limited information is available on the determination of the value of personal data. Tackling this from the perspective of company-related information (such as stock market valuation, revenues or profits per data record) yields some comparative indicators. Each method however has its own drawbacks in the precise calculation of the value of personal data. Adopting the perspective of the valuation by individuals themselves (behavioural economics), helps understanding some basic features on how individuals approach this valuation. Empirical studies on these features are scarce. Studies we discussed in this chapter indicate sensitivity of individuals for ownership of data, and the relevance of concepts such as instant gratification (evaluating immediate returns higher than returns on the longer term), hyperbolic discounting (difficulty of evaluating costs at the longer term against benefits at the shorter terms), and endowment effects (preferring the situation as it is beyond a new uncertain situation). The privacy paradox (people indicate they care about privacy but do not act accordingly) can be partly explained by reference to these behavioural features (but requires reference to options for choice offered and other societal features as well).

3 The PDM ecosystem

A Personal Data Market ecosystem is a system that aligns actors and relations between these actors in a systemic manner that itself is vital, robust and sustainable (i.e. has resilience towards short term and long term changes). Actors involved are the data subjects, i.e. the individuals whose personal data are collected, stored, aggregated, processed, disseminated and – in the end – destroyed or anonymised, the data service providers, i.e. the organisations that exert one, a few or all activities that can be performed in relation to the personal data and that has a direct link to the data subject, and third parties, i.e. the organisations that interact with the data service providers and that only indirectly interact with the data subjects while exerting one, a few or all activities that can be performed in relation to the personal data. The relations between the three main actors (data subject, data service provider and third party) can be unidirectional and bi-directional. They will be based upon trade-off principles that determine the added value (in economic and non-economic terms) of the relation. A vital PDM ecosystem implies that the relations between the dominant actors enable the exchange of data and the servicing of these data in a vital manner, i.e. in which returns on investments are sufficiently high for each of the partners to stay within the system. It is robust in the sense that the actors together cover a sufficiently large part of the possible exchange of personal data such that no additional partners are needed. And it is sustainable in the sense that some kind of equilibrium exists between the three types of actors, acknowledging each other's roles, motives and activities. In this section, we will translate the findings of the literature review into a conceptual model for a PDM ecosystem. We will start with an overview of the types of (personal) data involved, in order to be able to categorize different types of services according to the data on which the service is based. Furthermore, we will have a look at the perceived risks related to types of processing. Based on this, we will introduce a matrix which presents the different possibilities of incentives and control in a PDM.

Having described the types of data, services, and processing in a PDM, we will elaborate on the different key actors involved in a PDM, briefly introduced above.

3.1 Concepts for PDM ecosystems

We can look at different concepts to analyse the Personal data Market ecosystems. A first option is to look at the types of data that can be processed in a PDM ecosystem. Many collected data are personal data. The WEF (World Economic Forum, 2013) distinguishes three types of personal data:

- *Personally provided data*; data persons provide themselves, knowingly and willingly, for instance through postings, emails, filling in forms, etc. The fact that persons do 'personally' provide these data does not imply that they provide them voluntary, knowingly and willingly.. It might as well be involuntary (i.e. the consequence of having an account at Facebook), unknowingly (ruling out not knowing because of not having substantially checked what happens with one's data, but providing data that one really cannot know it will be provided), and unwillingly (no choice than to opt the delivery of data).

- *Observed data*: data that can be observed by parties that collect personally provided data and by third parties that do not collect personally provided data. An example of the first is an organisation with a website on which people search for information. The organisation can collect observed data for instance by collecting click data behaviour. An example of the second is an organisation (third party) that places cookies on a website to collect data about surf behaviour. Grey zones are present in this distinction: when a user accepts placement of cookies, should the data collected by these cookies be considered personally provided data or observed data?
- *Inferred data*: data collected either as personally provided data, observed data or data acquired through other means can be used to infer information about (groups of) persons. Even when a profile is constructed, it is possible that a direct linkage can be made to individuals. Profiles can have deductive 'capacities', implying that the profile is constructed and individuals are checked against these profiles, and they can have inductive 'capacities', meaning that the profile is created on the basis of input from various individuals.

The BCG report 'The Value of Our Identity' (BCG, 2012) uses a slightly different distinction between the various forms of data:

- Volunteered (Unstructured data)
- Required (Structured data)
- Tracked (Contextual data)
- Mined (Profiled data)

It uses a risk matrix in which each category of data is related to a specific use:

- Delivery of requested service
- Enhancement of service
- Transfer to third party (anonymised)
- Transfer to third party (traceable)

Through the matrix the report identifies which uses of data are most delicate for various domains, and require specific attention by the organisation. The colours are not so much related to legal and regulatory conditions but rather to presumed experiences and preferences by data subjects. An interesting issue is posed by the mined data: when this relates to aggregated data, it no longer is categorised as personal data and conditions from data protection regulation do not apply anymore (but other regulations such as the intellectual property rights still might apply!). Still, these are considered high-risk fields, since they may arouse potential unrest within communities.

Mined				
Tracked				
Required				
Volunteered				
	Delivery of service	Enhancement of service	Transfer (anonymised)	Transfer (traceable)

Figure 1: Example of a risk matrix as provided by BCG-report

Next to the types of personal data that are processed, the organisational structure and the incentives supporting a PDM ecosystem are extremely relevant. Whether the incentive is to monetize data or to protect the privacy interests of the individual makes an important difference. Similarly, it matters whether the data subject or the organisation (service provider) controls the data. The different options related to incentives and control are mapped in the matrix below.

		Organisation-centric	Subject-centric
Added value incentive	Financial reward	Organisation uses data (including selling to third parties) and gives money in return	Data subject sells data for specific purposes in return for money
	Non-financial reward	Organisation uses data to add value, such as delivery or improvement of services and personalisation	Data subject provides data in return for services or personalisation
Fundamental rights incentive (control)		Organisation applies PETs / PbD to protect rights of data subject	Data subject controls his own data via the platform of the service provider

Figure 2: Categorisation of Personal Data Markets in relation to incentive and control (TNO, 2014)

Organisation-centric means that the organisation is in control over which data will be collected for which purposes (within legal constraints). Subject-centric means it is the data subject who determines purpose and scope of data collection. Added value incentives refer to either the financial rewards that can be achieved by the data practices or non-financial rewards (such as improved service delivery, improved services, novel services, and personalisation of services).. Fundamental rights incentives refer to the protection or promotion of fundamental rights through application of privacy-preserving technologies, privacy by design, or through reorienting the control perspective over data collected and use so that the data subject is in control. When data are collected in a situation that can be characterised as organisation centric, it is the organisation determining scope and purpose of data collection. The organisation can have purely or strict monetary incentives to do so, strict non-monetary incentives or a blending of monetary and non-monetary incentives. An example of the first is ad revenues, an example of the second is the provision of traffic information on the basis of collected location data and an example of the third is a blending of the first and the second. It also can be focused merely on control of data and data flows. Similarly, subject-centric data collection may be oriented towards monetary valuation, non-monetary valuation and a blending of the two. It also could relate purely to control.

Within this matrix the four distinct fields enable categorising activities by an organisation or a data subject in the field of personal data collection and use. In our approach we presume to have one organisation being the key actor in the process of data collection and use, i.e. one key actor that determines purpose and scope of the data collection. We presume the data to be collected to contain personal data. It

may contain other data as well, but part of the data collected are personal data.[12] We presume the key actor may have connections to other parties (such as ad brokers, or third parties with which it has contractual agreements on delivering data for purposes not determined by the key actor). In the following subsections, the three key actors in a PDM ecosystem will be briefly elaborated upon.

3.2 The individual data subject

This is the individual to whom the data concern. Personal data are data relating to an identified or identifiable natural person. Identification can take place directly (e.g. by name) or indirectly (e.g. by an identification number). Personal data form the scope of this research. Data subjects play various roles within a PDM. Most importantly, they are the source of the personal data that are the key ingredients for the activities within the PDM ecosystem. They also serve as clients of services offered on the basis of the personal data delivered before. The personal data on which these services are based are either directly relating to the data subject in his/her role as client or indirectly relating to him/her. The service can be offered on an entirely personal basis, in which the person is approached *because* he/she is the person aimed at. It also can be offered because the person fits a profile that either is composed by using *inter alia* personal data coming from the data subject or that is applied to the data subject because personal data fit the profile of the data subject (that itself is created without using any personal data of the data subject). Data subjects can be related to each other in PDM ecosystems, and can exchange (personal) data with each other using infrastructures, networks and platforms offered by a service/content provider. An organisation as Facebook for instance, facilitates the exchange of (personal) data between data subjects who use the Facebook platform for the exchange of (personal) data. Facebook not only offers the platform that enables the exchange of data between data subjects, but also fulfils a role as data service provider.

3.3 The data service provider

The data service provider is the organisation that collects, processes, disseminates, stores, aggregates data stemming from data subjects and that offers services in exchange. The services not necessarily need to be oriented towards the data subjects who offer the data in first instance. The services may also be oriented towards other data service providers or third parties. The data service provider (DSP) has a direct link to the data subject through which the exchange of personal data is organised. The link between the DSP and the data subject can be a contractual relation, a service relation, or an incidental relation (‘just trying whether you are interested …’). A data service provider may facilitate the implementation of privacy-friendly constructions which can protect the privacy of the data subject, without hindering the commercial benefits of entities using the data.

A specific form of a data service provider is an entity that fulfils the role of identity broker. In this situation, three characteristics may be of guidance in assessing the

[12] The concept of personal data has been defined by directive 95/46/EC and the Proposal for the new data protection regulation as “any data relating to a data subject” (art 4(2) proposed Regulation). This is a broad definition, by the Art 29 WP to be understood as referring as well to data such as IP-addresses (see xxx) Opinion 2007/4 on the concept of personal data).

appropriateness of entities. These characteristics are: “(1) the identification process is *incidental* to the relevant firm’s overall business plan; (2) the initial and subsequent identification and verification processes are carried out *remotely*, and (3) the firm engaged in eID intermediation is operating in a lightly regulated setting.”(Zarsky & Andrade 2013, p. 1345)

The first characteristic, incidental to the core business, means that the entity is already processing personal data and is usually able to verify an individual’s identity, without this being the aim of the business. For instance, telecom providers or internet service providers have their clients for whom they provide a connection. Because of this connection and the underlying contracts, the providers can identify the individual related to a certain connection. Identification is, however, not the main goal, but is unavoidable when facilitating the service. Basically, to establish a connection between the individual (‘s device) and another device or a web server, it is necessary to know to which address or number the information needs to be sent. The core business is to provide a network service.

As regards the second characteristic, remotely, it is important that the individual does not have to show up physically to present himself. The identification process is automated and independent of a physical location of either party involved.

Thirdly, a certain level of a regulated setting is required. This regulated setting can be of help in providing a trustworthy relationship between parties. The legal backdrop provides a context and some rules to be abided. The regulated setting should be a setting which applies to the type of party involved at a more general level. So, a contractual relationship, regulated by contract law, is not sufficient in itself, even though it may be a trust-enhancing factor.

Other modalities of data service providers are DSPs that provision value added services, to individuals and to other DSPs, thereby creating a web of actors storing, exchanging, enriching, aggregating and disseminating (personal) data. These value added services can be of any kind, either directly attributable to a specific individual, or indirectly (through profile characteristics). The data service providers can use are so-called ‘personally provided’ data, ‘observed’ data and ‘inferred’ data. For an explanation of these terms, see section 3.1.

3.4 Third parties

The third party in this context is the entity that wants to make use of the personal data to provide a certain product or service, but without direct relationship to the data subject. This product or service can be provided to the individual, such as a personalized web environment or a social networking site. The provision is thus directed towards other entities. For instance, credit rating agencies process personal data to assess the creditworthiness of individuals. The results of these assessments are used by commercial companies who have to decide on whether they grant a loan or contract. In these cases, the individual may be affected, but is only involved as a source of data or as the object of an assessment.

Figure 3 shows the simplified eco-system with the key actors.[13] The main relation to be identified is the relation between the organisation that offers services and collects personal data for a variety of purposes, and the data subject who purchases the services on the basis of some kind of business proposition (paid services, freemium model, …).

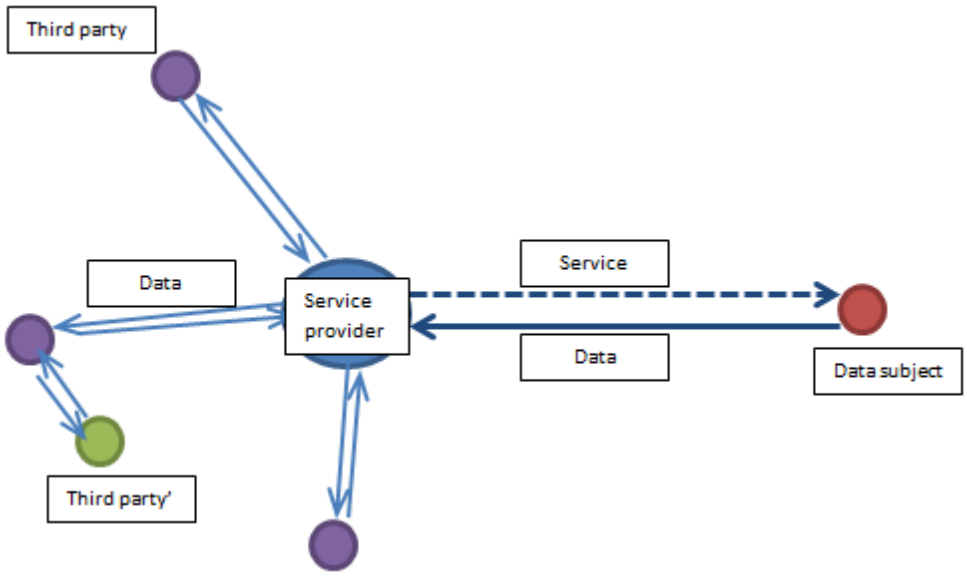

Figure 3: A simple, basic ecosystem with the service provider as an intermediary between the data subject and third parties

In the 'ecosystem' two roles are always present: the data subject and the service provider. Both can be the key actor, depending on whether the system is organization-centric or subject-centric. In our approach, the organization will have the role of service provider. The services provided can take different forms. In the case of services based on the incentive of creating added value, the organization uses the data to sell profiles or targeting services to third parties, or to deliver a service directed to the data subject. In case the services are based on a fundamental rights incentive, the service aims at providing control over the data to the data subject.

Obviously, third parties can take part in the ecosystem as well. The involvement of third parties depends on the service provided by the organization to the data subjects and the interrelation between the service provider and the data subject. In a situation where sensitive personal data will be exchanged a third party can act on behalf of the service provider. This is an approach found in the medical domain. Another role of the third party is to guarantee the trustworthiness of the organization (and of the data subjects) by the provision of certificates and by the provision of

[13] The concept of an ecosystem is a tricky one. Traditionally it refers to "a community of living organisms in conjunction with the non-living components of their environment, interacting as a system." (Wikipedia) Healthy ecosystems are considered to be sustainable, showing balance between the various components of the system. This sustainable aspect is implicitly part of the approach of data processing systems as ecosystems. It usually just refers to the set of actors and their relations in a specific data processing setting.

technical tools (especially cryptographic key management) that help organizing the trustworthy exchange of data between the service provider and the data subject (or between distinct service providers). This will be dealt with in the next section.

4 Use of personal data - some practical illustrations

Different types of services based on the processing of personal data can be distinguished. This section describes some of the more commonly existing types. While the direct exchange of data for money is rather uncommon, other direct types of exchange that one can observe in practice can be grouped in a number of categories: 'free services', with the aim of showing personalised advertisements, 'personalised offerings of products and services', with the aim of personalising the display of products on offer (and with the indirect aim of selling more services or products), and 'aggregated services', with the aim of selling personal data, but in an aggregated or anonymous[14] manner. These categories represent ideal types, in practice often hybrids exist. Furthermore, some companies have multiple business models at the same time. A last category that is separately mentioned is sensor and geographical data. The use of these data sources is quickly growing, and these data bear some interesting relations with personal data . Therefore, this category is mentioned separately. For describing the different categories, we specify the (most likely) legitimate ground for data processing, the most relevant use of the data, and the way data are acquired. Furthermore, we map every category to the table provided in Figure 2 to facilitate comparison.

4.1 'Free services'

Categorisation:

- Legitimate ground for data processing: consent, or legitimate interest of the data controller
- Use of data: showing advertisements
- Way of acquiring data: personally provided, observed and inferred data

This category can be mapped onto our classification of PDMs.

		Organisation-centric	Subject-centric
Added value incentive	Financial reward	Organisation uses data (including selling to third parties) and gives money in return	Data subject sells data for specific purposes in return for money
	Non-financial reward	Organisation uses data to add value, such as delivery or improvement of services and personalisation	Data subject provides data in return for services or personalisation
Fundamental rights incentive (control)		Organisation applies PETs / PbD to protect rights of data subject	Data subject controls his own data via the platform of the service provider

Figure 4: 'Free services' category related to actor centricity and incentive

[14] Once more, though we use the notion of anonymous data, in practice it becomes ever more difficult to keep personal data really anonymous. See for instance Opinion 05/2014 of the Art 29 Working Party on Anonymisation Techniques, adopted 10 April 2014, WP216.

Although it is also possible to deliver free services in a user centric manner, this is often not the case. Most common examples are organisation centric instead. Changes could be introduced to make these free services user centric.

Google, Facebook and many other web services use personal data as well as online behaviour of users to show specific content to specific users, most notably through display of advertisements. For companies aiming to attract consumers, this can be effective: the use of data to target specific consumers increased the response rate to an advertisement of Money U from 2% to 8%, for example.[15] Furthermore, through displaying advertisements, web services are able to provide 'free' services, such as maps, webmail, or social media. Inferred data based on either personally provided or observed personal data is used to provide internet users with advertisements that are suited to their preferences. Examples include the showing of advertisements based on a search term that is entered into a search engine (based on volunteered data), advertisements about products based on social media content provided (observed data), or the showing of advertisements (for example on news sites) based on the content of websites that were just visited (observed data).

4.2 'Personalised offerings of products and services'

Categorisation:

- Legitimate ground for data processing: consent, or legitimate interest of the data controller
- Use of data: personalized offerings of products or services
- Way of acquiring data: personally provided, observed or inferred data

This category can be mapped to our classification of PDMs provided in Figure 2.

		Organisation-centric	Subject-centric
Added value incentive	Financial reward	Organisation uses data (including selling to third parties) and gives money in return	Data subject sells data for specific purposes in return for money
	Non-financial reward	Organisation uses data to add value, such as delivery or improvement of services and personalisation	Data subject provides data in return for services or personalisation
Fundamental rights incentive (control)		Organisation applies PETs / PbD to protect rights of data subject	Data subject controls his own data via the platform of the service provider

Figure 5: 'Personalised offerings of products and services' category related to actor centricity and incentive

Although it is also possible to deliver personalised services in a user centric manner, this is often not the case. Most common examples are organisation centric. Customers do not have a choice whether to provide data or not. Changes could be introduced to make these services user centric.

[15] Based on an interview with Robert Feltzer, EDM.

Many web shops target visitors with ‘personalised’ offerings with the aim of selling more products or services. Well-known examples include Amazon and Bol.com, selling books by making offerings, stating: ‘people that ordered this, also liked that’. This can be done using collected data, for example based on earlier visits or purchasing history of customers. Amazon, for example, is a company that uses big data to improve its automated recommendation system (see inset).

Example: Amazon
Amazon is the world’s largest online retailer. While it started out as online bookstore, the company quickly diversified, selling DVDs, CDs, MP3 downloads/streaming, software, video games, electronics, apparel, furniture, food, toys, and even jewellery. The company possesses personal information, purchase patterns and preferences from more than 152 million customers. These data are used for multiple purposes, such as tracking Amazon’s products throughout the supply process, but also offering advanced services to their customers and approaching each customer as an individual. Amazon has a clear understanding of what people actually buy and is able to make offerings for what else people may want. Furthermore, Amazon uses the data it collects to create ‘aggregated services’ (see further on), to deliver targeted advertising on third-party sites across the web, to sell data of its customers (in batches) to third-parties such as marketers or brands, and to offer suppliers options to promote products/categories on their own website to increase visibility with customers.

Amazon is able to combine historical client data with behavioural data (‘click data’) of customers on their website. They use the data gathered mainly for making better commercial offers to their customers, but at the same time they are selling data on customer profiles to others (see category of ‘aggregated services’). A second example in this category is Netflix (see inset). Netflix is an example of personalization through data, as Joris Evers, the director of Global Communications at Netflix[16] said: “There are 33 million different versions of Netflix.”

[16] Carr, D. (2013) Giving Viewers What They Want. In: New York Times. Available via: http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html?_r=0 .

Example: Netflix
Netflix is an internet video streaming service offering an all-you-can-view subscription to its collection of over 100.000 movie and television titles. One of the main applications of big data analytics at Netflix is its personalized recommendation system, leveraging billions of hours of subscribers viewing data. The goal of this data-driven innovation is to refine a seamless user experience to promote member retention. Additionally, it uses data analytics to determine what new titles Netflix should acquire for what kind of price – and even to commission the production of its own original content. Furthermore, Netflix uses data analytics to improve its marketing efforts and to handle internal production and processing activities. The data that Netflix uses, is largely collected in its own Netflix environment, generated by its 33 million users based on their behaviour, ratings et cetera. Furthermore, it acquires data from social media such as Facebook, and data from content providers, ISP's and financial service providers. Most user data is collected via the Netflix environment that is available via devices. The data that is being collected about members is both generated manually and automatically. When new member set up an account, they are asked to name a few of their favourite films and TV series. Furthermore, they can rate films and series. In some cases Netflix also uses data from external sources that may be generated deliberately in the original context but not with the intention that these data would be used by Netflix.

4.3 'Aggregated services'

Categorisation:

- Legitimate ground for data processing: consent or legitimate interest of the data controller
- Use of data: developing new services
- Way of acquiring data: inferred data

This category can again be mapped to our classification of PDM's.

		Organisation-centric	Subject-centric
Added value incentive	Financial reward	Organisation uses data (including selling to third parties) and gives money in return	Data subject sells data for specific purposes in return for money
	Non-financial reward	Organisation uses data to add value, such as delivery or improvement of services and personalisation	Data subject provides data in return for services or personalisation
Fundamental rights incentive (control)		Organisation applies PETs / PbD to protect rights of data subject	Data subject controls his own data via the platform of the service provider

Figure 6: 'Aggregated services' category related to actor centricity and incentive.

As Figure 6 shows, 'Aggregated services' does not map to any of the categories, because strictly speaking this is not a form of PDM as personal data are not processed and no reward is given to individuals.

Many organisations process personal data because they need this information to deliver their service(s) to their customers. Examples include energy companies that need the energy use and the address of their customers for billing them. Based on these data, some organizations develop new products and services that they sell on. Google is a company that has done this for several activities it is engaged with. The company, for example, developed Hadoop for internal purposes but then realized that others may want to use it as well and turned it into a commercial product. Amazon, as described above, has also done the same with the development of services based on its wealth of consumer profiles. A third example is Equens (see inset). Turning data which is not personal data into new services, as in the example of Hadoop, does not require a legitimate ground for the processing as no personal data are involved. Turning personal data into aggregated data that can no longer be traced to individuals, as is the case in the example of Amazon still requires consent, or a legitimate interest of the data controller. The third example, Equens, appears similar to the example of Amazon, but is actually more related to the Hadoop example. Equens does transfer personal data through its systems, but nowhere actually accesses this data. The only data they use is non-personal data on the traffic flow of these data, not the data itself. Using these anonymous flow data to develop services, means that no personal data is used and a legitimate ground for processing is not necessary.

Use case: Equens
Equens is an entity set up by the banking sector to process electronic payments (the PIN payment system) in the Netherlands. They are currently exploring whether new services can be developed, based on these PIN transaction data. Examples of these services are the analysis of transactions for specific stores in order to provide insight in the amount of returning customers, and the percentage of customers which generates the major part of the turnover. Equens has been in the news in a very negative way, because of alleged privacy violations. They were even nominated for a Big Brother Award by Bits of Freedom. However, Equens is not a controller in relation to the processing of personal data, but only provides ‘traffic management’ on behalf of the banks. As described above, they only aim to use the traffic data for developing new services. This means that based on this traffic data they are able to infer whether a customer is returning or not, without having to actually process any personal data. The service could be compared to anonymous and aggregated analytics, without processing of any data at an individual level. Still, for many people making this distinction is difficult, and it is therefore interesting that their case received so much negative publicity.

4.4 Geographical and sensor data

Categorisation:

- Legitimate ground for processing: consent or legitimate interest of the data controller
- Use of data: personal data for further processing and services.
- Way of acquiring data: observed and inferred data

		Organisation-centric	Subject-centric
Added value incentive	Financial reward	Organisation uses data (including selling to third parties) and gives money in return	Data subject sells data for specific purposes in return for money
	Non-financial reward	Organisation uses data to add value, such as delivery or improvement of services and personalisation	Data subject provides data in return for services or personalisation
Fundamental rights incentive (control)		Organisation applies PETs / PbD to protect rights of data subject	Data subject controls his own data via the platform of the service provider

Figure 7: 'Geographical and sensor data' category related to actor centricity and incentive

Similar to the first two categories, these services could be made user centric as well as organization centric. There are few examples of this category, but those that are found are organization centric.

Geographical and sensor data represent a slightly different category. These data may or may not represent personal data, depending on whether they can be traced back to individuals. A link to an individual is usually made by connecting a mobile device of a person to the location. This can be dynamic, such as the location data of a smart phone or the location data of a navigation system that is fixed to a car, or static, such as a check in portal for public transport, which makes the connection to an individual when he checks in or out, but does not monitor movement continuously between start and end point.. Geographical and sensor data can be are enriched with other data sources, and they can be combined with observed or inferred data. Often, geographical and sensor data are aggregated to create geo-services that organisations can sell or offer to other parties. The aggregated data are, if properly anonymised, (no longer representing personal data). Examples of this use include the use of telephony data of Vodafone by Mesuro to do analyses on traffic data. A comparable service is TomTom, who use real-time traffic (movement) data of its users to update its services, for instance to produce congestion information ('crowd-sensing'). Geographical data is included in the new EU Regulation on data protection as these data can increasingly be characterised as personal data. When you can map someone travelling five days a week from the same home address to the same place, it becomes very easy to track someone down. Sense-OS, a company specialized in streaming sensor data and building platforms for interpretation and contextualization of sensor data adopts the position that the only way to use these data in a privacy friendly manner, is to make sure that consumers retain control over their data via 'opt-in' mechanisms.[17]

[17] Interview with Jan Peter Larsen.

5 Use cases

PDM ecosystems can take different forms. The functionality and related services and incentives can also differ, mainly depending on the context. In this section, three use cases will be described. First, an individual PDM, which was the initiative of a Dutch student who sold his own personal data in an auction. Second, a use case in the field of health care . And third, a use case in the financial sector, where recently quite a lot of attention has been paid to banks reusing personal data of their clients.

5.1 An individual PDM: BuyMeOut

The value of personal data is clear, as well as the financial gain made by commercial companies who collect and sell personal data. One of the incentives that may form the foundation for a personal data market is to have the individual gain as well. An organization may act as an intermediary to facilitate this for a large quantity of people. However, in April 2014, a Dutch student sold his own data directly via an online auction.

This use case is mapped onto our classification of PDMs in Figure 8.

		Organisation-centric	Subject-centric
Added value incentive	Financial reward	Organisation uses data (including selling to third parties) and gives money in return	Data subject sells data for specific purposes in return for money
	Non-financial reward	Organisation uses data to add value, such as delivery or improvement of services and personalisation	Data subject provides data in return for services or personalisation
Fundamental rights incentive (control)		Organisation applies PETs / PbD to protect rights of data subject	Data subject controls his own data via the platform of the service provider

Figure 8: ‘Individual PDM’ category related to actor centricity and incentive

The student, Shawn Buckles, started the experiment in order to make a statement. Exactly for the reason described above, he wanted to receive money for his personal data. In an online auction he offered his data for sale and described what data would be obtained by the highest bidder. These data included:

- Personal profile info
- Location track records
- Train track records
- Personal calendar
- Email conversations
- Online conversations
- Thoughts
- Consumer preferences
- Browsing history

The data to be sold contained all these data over the period from April 12, 2014 to April 12, 2015. The buyer would, thus, be sure to receive all data for the next year. This also ensures that the data would be up-to-date.

The complexity of the experiment became clear at an early stage. Mainly from a legal perspective, several barriers appeared to be present. For instance, when selling a chat history, messages from others with whom you have a chat conversation are included as well. These data should be excluded from the data to be sold. Alternatively, consent should be obtained from the others for the selling of their data. In the experiment, only data from people who consented is included. All data from those that did not explicitly consent is deleted.

Another important issue is the fact that the data are not unique. All data are already collected and further processed by several companies. For instance, ISPs and telecom operators facilitate the services for browsing and calling and process the data on browsing history and phone calls. The data on train tracks are processed by the Dutch railways and the public transport chip card provider.

The experiment is meant as an awareness action. The money received from the auction will be donated to Bits of Freedom, a Dutch digital rights movement.
It is clear that the action is not (endlessly) repeatable, since the benefits will depend on the uniqueness of the action. If numerous people start offering their data for sale in this manner, the value will be limited. To compare, in ordinary situations of data selling, the value is in the mass of people to be targeted. The value of data about one individual is extremely low.

In total, almost 50 bids were brought out, of which the highest bid was 350 euros. The data were sold for this amount to The Next Web (TNW). TNW will use the data for an awareness action at a conference, where they will show how much they know about a person from the audience (Shawn Buckles) based on the data. They will single him out and show the audience what they can do with the data. Because of this application of the data, TNW does not have to receive all data that were offered in the auction. Instead, they will receive a number of screenshots with data and information which they can use. This allows Shawn Buckles to remove data from others when necessary. The result is, thus, more privacy friendly than originally expected.

It is remarkable that the action caught attention from the Dutch Data Protection Authority and that this authority indicated that selling data from others is prohibited if they are not collected for this specific purpose. The whole idea is to provide data from private interactions, so the data are initially collected for private purposes. When data from others is involved, this data is processed for private purposes and not for the purpose of selling the data. This even seems to be problematic when explicit consent has been obtained from the relevant data subjects. Here, the DPA makes a distinction between private persons and commercial entities. Once you act as a private person, the data are most likely processed for private purposes as well. This reasoning is in line with the idea of the household exemption, which allows for processing of personal data from others for ordinary, non-commercial private purposes.

Selling personal data from one individual person without the use of an intermediary is an atypical example of a personal data market. The value attributed to the data in the above example, 350 euros, cannot be seen as a representative amount. The uniqueness of the action made the data attractive and the highest bid came from an organization with a similar idea of awareness creation. The data will, thus, not be used for commercial purposes by the receiving entity. The case made also clear that selling personal data is quite complex. Even when explicit consent has been obtained from other individuals whose data is included, for instance, because their contact details are in the telephone book or because they are participating in a chat conversation, it may not be allowed to sell these data. Also in this case, consent has to be obtained beforehand and for the specific purpose of collecting and selling the data. If the system of selling personal data at an individual level would work in practice, the result would be that individuals might start asking for a share from others who sell data which include their data as well.

The student wants to extend his action by a next step. What this step will be is still unclear, however. His wish is to offer alternatives for current applications and services which allow for selling your own data. Nevertheless, these alternatives seem to be scarce and often technically complex. Moreover, exclusivity of the data cannot be guaranteed, so there are always other entities who receive and process at least parts of the data involved.

5.2 Towards a PDM ecosystem in health care

5.2.1 Introduction

The Dutch health care system is a complex system. A multitude of practitioners contribute to curing of and caring for the patient, both within (para-)medical institutes (intramural) and outside (para-) medical institutes (extramural). The Netherlands is one of the few European countries that have adopted a market approach to health. The three main actors, the patients, the insurers and the care providers, compete on resources. Insurers compete on patients, care providers compete on insurers and patients are able to optimise their search for care quality to a certain extent. The medical industry (pharmaceutical organisations, organisations offering medical equipment) will need to have competitive offers to be selected by care providers. Admittedly this is a very simplified scheme. One crucial element that needs to be inserted is the role of government as regulator. Notwithstanding some competitive elements, the Dutch health market is a very regulated market, in which several regulatory constraints are built in that hamper the functioning of a real free market.[18] Privacy and personal data protection form part of the regulatory constraints.

Privacy of the patient is key to medical treatment. Medical data are considered sensitive data for which more stringent data protection regimes are in place. With the advent of the ‘big data revolution’, medical data become more interesting for a variety of stakeholders since they may embed interesting information on several issues related to health care, such as information on co-morbidity schemes, on

[18] See for instance ‘Wijziging van de Wet marktordening gezondheidszorg en enkele andere wetten, teneinde te voorkomen dat zorgverzekeraars zelf zorg verlenen of zorg laten aanbieden door zorgaanbieders waarin zij zelf zeggenschap hebben’ in which Dutch government discusses its strategy regarding the market organisation of the Dutch care system. (TK 33 362; 1 July 2013).

insurance practices, on quality of medical services, on quality of medical institutions, etc. Having the medical sector as one of the fastest growing economic sectors in the world[19] and the data intensity of the medical sector increasing likewise (to say the least), we expect the approach of the medical sector in terms of a personal data market to offer interesting novel perspectives. The use case in the health sector is mapped onto the classification of PDMs in Figure 9. As most use cases deal with sensitive data, they protect the personal data using PETs and PbD. The aim is to achieve benefits in the form of better insights and research for health purposes in general, so the benefits are not directed towards the specific individuals whose data it concerns.

		Organisation-centric	Subject-centric
Added value incentive	Financial reward	Organisation uses data (including selling to third parties) and gives money in return	Data subject sells data for specific purposes in return for money
	Non-financial reward	Organisation uses data to add value, such as delivery or improvement of services and personalisation	Data subject provides data in return for services or personalisation
Fundamental rights incentive (control)		Organisation applies PETs / PbD to protect rights of data subject	Data subject controls his own data via the platform of the service provider

Figure 9: 'PDM system in healthcare' category related to actor centricity and incentive

5.2.2 *The health care system as a personal data market*

The exchange of personal data within the healthcare system is bound to a number of strict conditions. Medical treatment requires consent of the patient. Patient data are considered to be sensitive data that need strict technical and organisational measures to protect the sensitive relation between practitioner and patient and to prevent abuse and misuse of these data by third parties. Next to the national Data Protection Directive, the Regulation for Medical Treatment (Wet op de Geneeskundige Behandelingsovereenkomst) serves as the regulatory backbone to decide what can be done with gathered medical data.

In the present situation the role of the individual patient in deciding what will be done with his or her data is limited. People do worry about the potential abuse of medical data, as surveys in this field show.[20] Loss of medical data, data breaches or criminal use of medical data are a few of the worries. On the other hand, the use of personal data for improving medical treatment is widely acknowledged as worthwhile.[21] To give one example, notwithstanding the very positive attitude of a Patients-like-me-participants with respect to the use of personal medical data (94% supportive for using personal data for improving medical treatment), the majority of

[19] See http://www.triple-tree.com/research/healthcare/
[20] http://www.computerweekly.com/news/2240106420/Fears-over-medical-record-privacy-could-deter-patients-seeking-treatment-finds-survey
[21] http://news.patientslikeme.com/press-release/patientslikeme-survey-shows-vast-majority-people-health-conditions-are-willing-share-t. The survey was performed under over 2000 participants to patients like me. This is a biased sample, having patients who most probably will profit from sharing data on rare diseases.

respondents also raised concern for potential exclusion of medical treatments because of their illness and exclusion of jobs!

In the Netherlands a number of institutions deal with the collection and dissemination of health care data. The three most prominent ones are: TI Pharma, Mondriaan and Parelsnoer. Of these three, Top Institute Pharma embeds the largest set of institutes: 45 (global) business partners and 28 knowledge institutes. TI Pharma is a not-for-profit organisation, and presents itself as a pharmaceutical enabler.[22] It runs twelve scientific programmes that focus on specific diseases and on pharmaceutical research activities that help improving drugs delivery and effectiveness. The Board of Directors consists of industry and academic representatives. TI Pharma aims at realising public-private partnerships between the pharmaceutical industry, academia and health care practitioners. One of its projects is the Mondriaan project, part of the theme 'Efficiency Analysis of the Process of Drug Discovery, Development and Utilization'. Mondriaan essentially is a mediator between data owners, on demand bringing data together of various data sources.

The LifeLines project is a relevant project in studying factors that may help in preventing diseases. The LifeLines project started in 2006 with a longitudinal research effort in which 165.000 patients in Northern Netherland (provinces of Groningen, Friesland and Drenthe) are followed over a period of 30 years. The sample includes three generations of people. Participants are invited for a medical check every five years and are screened on a number of issues (blood, genetic aspects, health conditions). LifeLines is hosted by the UMCG and has developed into an expertise centre on biobanking and cohort analysis. Cooperation with national and international research institutes is actively pursued. The UMCG is connected to Parelsnoer.

In Figure 10 we have depicted the ecosystem of dominant players in the Dutch health care market for health care data. We only included one insurer, Achmea, as illustration for the role of insurers. Achmea is an interesting player since it houses the Achmea Health Database, the successor of the former Agis Health Database, with reimbursement data of care consumption of 4.7 million Dutch individuals. Achmea relates to other key players active in health care data. For Achmea, the Achmea Health Database is a relevant repository of reimbursement data of care consumptionthat can be used for a variety of purposes (see below). For external use it only opens up its data for scientific purposes.

The upper half of Figure 10 shows the initiatives that are within the public domain, and specifically relate to scientific activities. The lower half includes private activities, and relates to commercial products as well.

[22] See http://www.tipharma.com/ (visited April 24 2014)

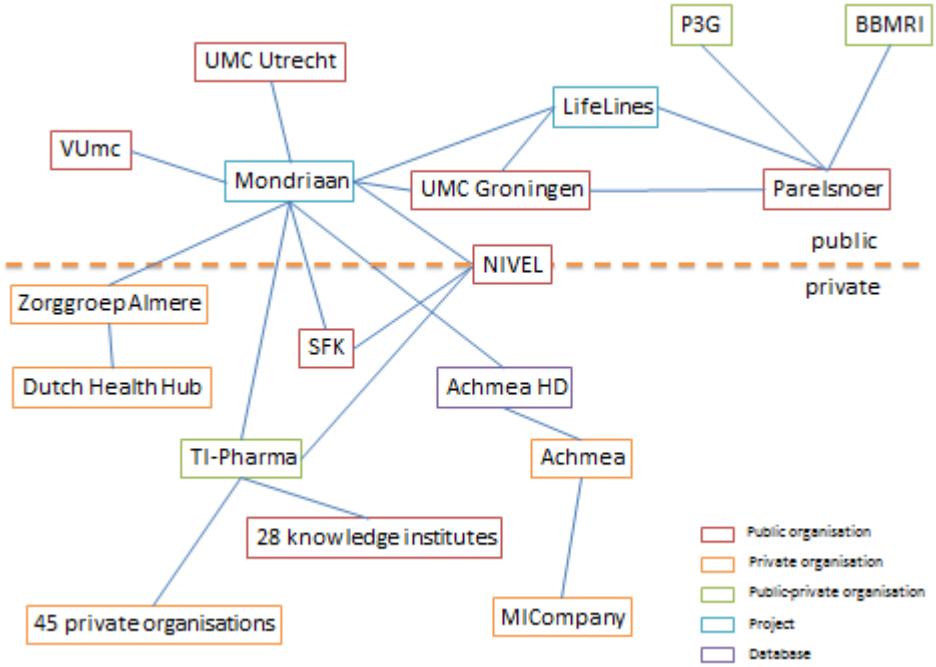

Figure 10: Simplified ecosystem of data gathering by health organisations, the Netherlands (TNO 2014)

Data-crunchers such as Stichting Farmaceutische Kengetallen (SFK) and NIVEL form data backbones of medical practices in the Netherlands. SFK collects data on prescription of medicines by public and hospital pharmacies. Its purpose is to promote good practices of medicine prescription, to promote scientific practices and to serve the interests of the pharmacists.[23] It represents over 95% of Dutch public pharmacists and collects data of over 15,3 million persons.[24] It gathers data to be delivered to third parties, who in turn need to submit a motivation for using these data (and need to pay a suitable remuneration). NIVEL is the national hub for first line health care and collects data through the LINH of 386 general practices, which comprises 1,2 million patients; some 540 primary psychologists with annually 46.000 patients; some 60 physiotherapists with annually 5.000 patients; some 50 practices Cesar and Mensendieck with annually 4.000 patients; some 60 dietary practices with annually 5.300 patients; and some 130 pharmacies with annually 20 M prescriptions (which it receives through SFK).[25] NIVEL has a number of research programmes in which it investigates quality of and potential improvements in medical practices. These programmes relate to direct medical interventions (home care for instance) and the organisation of care processes.[26]

An interesting initiative with health care data is the Almere Data Capital (ADC) which is 'under construction' in Almere city. One of the pillars of the ADC is the Dutch Health Hub, that intends to connect medical research departments (UMCs), industry, medical practitioners, and quality assurance institutes in health. The Dutch

[23] See http://www.sfk.nl/pdf-documenten/sfk-algemeen/statute (visited April 23, 2014)
[24] See http://www.sfk.nl/over-de-sfk (visited April 23, 2014)
[25] See http://www.nivel.nl/NZR/zorgregistraties-eerstelijn (visited April 23, 2014)
[26] See http://www.nivel.nl/ (visited April 23, 2014)

Health Hub presents itself as a big data services platform, offering services to health care providers, industry and government organisations. It is an association with members. At present, mostly IT suppliers participate. The Board consists of representatives of IT service organisations, active in the field of health information services and health applications (Capgemini, HP, IBM, KPN, SurfSara, Vancis, UNET as large organisations; Almere Grid, 17Rabbits, 22 times, Carelliance, MijnEGD, Hospi-Trace, PS-Tech, Semantoya and Vicinitas as involved SMEs).[27] ADC intends to play a role in offering commercial services in organising health information processes. It aims to develop three Proof of Concepts: Digitooth (on dentristy), a Centre for Genetic Diagnostics and a national reference centre for population screening and research (initially focusing on breast cancer). For all PoC, business cases need to be developed that show viability of these initiatives. Zorggroep Almere is one of the participants in ADC.

5.2.3 Case study: the value of personal data in health care

As an illustration we looked in more detail at the activities of four major players in the Dutch ecosystem: Mondriaan, being a project with a clear link to an institution of public and private actors (TI Pharma), Achmea, already acting as a private actor and being active in the market of personal data through its Achmea Health Database, and LifeLines as an initiative gathering data of individuals during different phases of their life. We studied their relation with big data developments and privacy aspects by means of desk research, document analysis and an interview with representatives of the institutes.[28] Table 2 presents an overview of the main results.

Table 2: Comparative overview of main features of Personal Data Markets in health care (TNO, 2014)

	Achmea	**Mondriaan**	**LifeLines**
Structure	Company	Foundation	Project
Size	50 persons in knowledge centre	4 persons	130 persons (90fte)
Aim	Improve insurance products Improve medical treatment	"Create a grid to integrate and to enrich existing and new health data platforms …"	Collecting data on life courses of 165.000 persons over a time span of 30 years; data collected relate to life style and quality of life ('healthy ageing')
Data collection	Achmea Health Database	No data collection; having data sources combined on the basis of specific requests by an independent TTP itself	Surveys (every 1.5 year) visits (every 5 year) Biomaterial
Size of data collection	4.7 million patients Xx Healthhealth practitioners Xx health institutes	Associated partners: SFK (>90% of pharmacies; 14M patients); AHD (4.7 M patients); Zorggroep Almere (GP	165.000 persons of provinces of Groningen, Friesland and Drenthe

[27] See http://www.dutchhealthhub.nl/bestuur-en-leden (visited April 23, 2014)

[28] We are indebted to Willem de Bruin (Mondriaan), Barry Egberts (Achmea) and Marie-José Bonthuis (LifeLines) for their participation to the interviews.

	Achmea	Mondriaan	LifeLines
		data of > 200.000 patients) NIVEL UMC Utrecht (GP data of > 250.000 patient) VUmc (GP dat of > 200.000 patients) UMC Groningen (Lifelines 40.000 patients) Psychiatric Case Registry (data from 600.000 patients from psychiatric institutions	
Data processing	Self Through TTP (ZorgTTP)	Through an infrastructure which is able to pseudonymise, transport and link data. This infrastructure also includes a TTP (Custodix)	UMCG (TCC)
# requests by external parties	~20/year	~4/month	~200/year
Kind of requests	Mostly UMCs, universities/RTOs,; few from pharmaceutical industry; mostly scientific research proposals		Mostly UMCs, universities/RTOs,; few from pharmaceutical industry; mostly scientific research proposals
Internal privacy procedure	Privacy commission Privacy officer	Privacy impact assessment (quick scan)	Privacy officer (FG)
Conditions	Results must be made publicly available	Sources have to agree to disclosure of their data to a requesting party	Results must be made publicly available
Data protection measures	Pseudonymisation according to guidelines of CBP (Dutch DPA)	Several Privacy Enhancing Techniques and Procedures to anonymise data, but also to prevent loss of anonymity downstream in the data-chain	Pseudonymisation according to guidelines of CBP (Dutch DPA)

5.2.3.1 Achmea

Achmea is a large organisation that offers, among others, health insurances. It has organised its strategic health care activities in a separate unit, called the knowledge centre, which falls directly under the director of the division Health and Care. This knowledge centre has a total of 50 persons working on business intelligence, data warehousing, research into quality of health care, and business consultants oriented towards data driven, fact-based new business methods. Achmea uses a 'need-to-know' approach for access management to systems containing personal

data of clients (and personnel). Personal data are, anonymized, increasingly used for quality improvements, for measuring effectiveness of care, for customer profiling and for forensics (fraud detection). Achmea has one of the largest health databases in the Netherlands with data on 4,7 million insured and health practitioners.

For scientific research Achmea receives about 20 requests per year for access to its data, mostly by academic institutes. It assesses requests on the relevance of the research question, the appropriateness of the approach chosen and the assessment whether requested data can indeed help answering the research question. All results must be made public in peer reviewed journals, but before making them public, Achmea performs a check to assure the scientific soundness of the publication. It has asked an external organisation to perform an audit of the entire process of data processing on a regular basis, to be sure to remain on the safe side. For external purposes for linking with external databases, Achmea uses ZorgTTP that follows the guidelines of the Dutch DPA in pseudonymising the data (see box). Since this is accepted as anonymising, patient consent is not strictly necessary anymore.

Achmea senses the tension between on the one hand the societal mission of an insurer to deliver the best products at the best competitive pricing and on the other hand the societal reluctance that all conceivable means will be used to realise this (i.e. by using all data on a personal level). It perceives the risk of a (real or alleged) privacy infringement and the subsequent impact this has on the reputation of the organisation. As an example, recently it faced a patient who refused to have its disease indicated on his billing. But without this information an insurer iis not able to check if the treatment is legit. This kind of sensitivities will grow and require a proactive attitude of the insurer.

If alternative models enable to meet the conditions which patients consider relevant and Achmea still will be able to fulfil its primary role, than Achmea would not oppose this. It perceives the relevance of the discussion on privacy protective measures, and it certainly supports technical solutions that may help solving privacy threats.

5.2.3.2 Mondriaan

A consortium of 12 organisations initiated the idea of Mondriaan, a project that should enable the linking of available data sources with medical data in the Netherlands for specific purposes. Being part of TI Pharma the focus of the project is on improving medical treatment and quality/effectiveness of care. Mondriaan only recently started. Essentially, Mondriaan is a data broker, bringing together medical data of different sources depending on the specificities of the request. It operates as a non-profit service provider. It does not collect data itself but seeks that data of different sources are gathered and made available. To this end it uses a infrastructure which extracts and pseudonymises data from sources; and links data from different sources on subject level. Mondriaan uses a TTP, Custodix, for the pseudonymisation and record linkage as well as to enable separate processing of communication (identifiers) and research data. Mondriaan acknowledges that, notwithstanding following the guidelines of CBP, the Dutch DPA, this technique may not be sufficient to offer full anonymisation in all cases. Identification of single patients cannot be excluded when (pseudonymised) data of different sources are combined in sufficient detail. Mondriaan expects that informed and explicit consent will play a larger role in the near future to enable using medical data. Up till now, Mondriaan receives about three to five requests per month for linking data sources. Mondriaan but also the sources assess the relevance of the request on scientific

grounds, the appropriateness of the requested data for these purposes and the quality of the approach. Mondriaan and the sources do not cooperate when commercial purposes (such as marketing) are the primary goal of the request. So far, mainly academic institutions have approached Mondriaan, but it is expected that – given the objectives and composition of TI-Pharma – pharmaceutical industries may become more interested over time. Mondriaan uses an in-house developed Privacy Impact Assessment to assess the privacy aspects of a request. This PIA is a kind of quick scan, enabling a quick assessment.

5.2.3.3 LifeLines

LifeLines is a research project that started in 2006 with the aim to collect data on a large number of persons in the three Northern provinces of the Netherlands over an extended period of 30 years. Data collected relate to socio-demographic features, quality of life characteristics and issues related to healthy ageing. Data is collected through regular surveys (once every year and a half) and visits (once every five year). During visits, biomedical material is collected as well. A total of 165.000 persons living in the Northern part of the Netherlands at the start of the project are involved. Since February 2014, Lifelines is an independent organisation with the Medical department of the University of Groningen (UMCG) being the full owner of LifeLines. LifeLines employs about 130 persons (90 fte). It is financed by the Ministry of Health which has enabled the project to follow 165.000 persons in the period 2007-2017. It cooperates with other non-commercial organisations that collect personal data such as Bioshare, the BBMRI and NFU.

LifeLines receives about 200 requests per year from scientific researchers for access to its data. Lifelines operates on a non-commercial basis and only charges marginal costs for providing the data. It only provides access for non-commercial purposes. Requests are assessed against criteria of scientific relevance and validity/feasibility of the research question. LifeLines uses a TTP for using pseudonymisation techniques, following CBP guidelines. On top if this, LifeLines investigates with the University of Amsterdam the options to include k-anonimity (meaning that no individual person can be singled out when in a group of at least k individuals). Involved persons have given explicit and informed consent to the collection of the personal data. Use of data has been described in a generic though sufficiently specific manner so that no consent is needed for individual research requests. Only when new uses come into play, new consent is asked..

LifeLines has a privacy officer (Functionaris gegevensbescherming) and investigates at present the need for implementing a Privacy Impact Assessment that should be applied for internal purposes and to assess requests for data delivery. Data subjects are informed about their rights regarding access to their data, withdrawal of consent, and complete removal of their data if they wish. LifeLines considers transparency towards data subjects essential. Though the upcoming new EU Regulation on Data Protection) will strengthen privacy requirements, it will not stop (nor is it the intention to do so) the on-going commercialisation of personal data. This will require continuous political attention and awareness.

5.2.4 A business approach towards privacy and big data in health care

One of the initiatives in the Netherlands concerning the storage and handling of personal medical data is the Almere Data Centre. The ADC explores whether it is

feasible to act as a hub gathering medical data and acting as a data broker. The ADC intends to create a market for personal medical data. Privacy issues are prominent in this role. The ADC consists mainly of service and solution providers. The ADC commissioned a report to investigate issues concerning privacy and data protection (Neuteboom, 2012). The report details issues concerning data ownership, ownership of records, and investigates the role of a third party that acts as processor (the role of ADC). The conclusions of the report, namely that ADC could act lawfully as central warehouse of medical data, was criticized in another publication on a number of issues (see below; Veen 2012).

The discussion shows the sensitivity of the issue and reveals a number of concerns re. privacy/data protection. Starting point of the ADC is the central storage of medical data from a large variety of sources, in order to help improving medical research and quality of care. Though these goals are not disputed, a number of critical issues remain:

- Basic to all data processing is the proportionality, the necessity and the subsidiarity of the processing. Proportionality indicates that the use is not excessive to the goal, necessity means it has a lawful or socially valid reason, and subsidiarity means that no other means are available that could achieve a similar objective at less (social, organizational, financial) costs. When centralizing medical data, these principles need to be addressed and need to be affirmatively answered. The critical report questions whether this has been sufficiently addressed in the ADC case, and whether it is sufficiently in the minds of those trying to develop the ADC business case.
- A difficult to determine legal issue is the responsibility of the ADC, especially whether it acts as a controller or as a processor. Given the aim of the ADC, namely to collect medical data of a variety of sources and use them to improve health care, most probably the ADC will be identified as a controller. This implies it has its own responsibilities in meeting the constraints of the data protection regulation in place. A relevant aspect of this responsibility is whether data collected for one purpose may be used for another. Especially the objective of ADC to ‘commercialize’ data collected, may be problematic since this may be too far out of reach of the original intention of the collection of data. The fact that data will be anonymized is not relevant, since the purpose of commercialization is related to the personal data collected in first instance, as the critical report indicates.
- Another issue raised by Veen is the potential use of privacy enhancing technologies (privacy/data protection by design) and privacy impact assessments in order to act responsibly and to take all necessary precautionary means to prevent data leakages, to identify privacy risks and to promote safe and responsible use of personal medical data throughout the organization. This helps in promoting data security and safeguard privacy of patients. Veen questions whether ADC is sufficiently aware of these opportunities.
- Next, ownership of data is an interesting feature to be investigated in depth. Both reports acknowledge that ownership of personal data sec does not exist (is not acknowledged to be a juridical entity), while ownership of the material substrate carrying the data does exist (such as the paper on which the personal data are registered). This has led to intriguing approaches by Dutch courts in deciding in specific cases who owned what precisely. With data stored at servers at the ADC, the issue of ownership needs further examination. The

outcome of this examination may be relevant for the determination of the value of the data stored at these servers.

While both reports give better insights into the relevance of an appropriate approach towards privacy preserving features and awareness built in in ambitious initiatives as ADC's health hub, it stays away from the central question of our study, namely whether a market for personal data market will develop and how this should look like. The approach we sketched in chapter 3 relates to the valorization of personal data. When assessing the ADC, the purpose is to valorize personal data but with limited inference of the patients themselves. The focus is on aggregated data, in which pseudonymisation and anonymisation techniques are used. Still, given concerns on privacy with large data sets that may lead to re-identification, both reports identify several issues that should be taken into account. Further study is needed to determine the regulatory approaches that need to be in place for the proper treatment of personal data. The two reports show that several issues are not satisfactorily solved at present. Both reports give motives for further consideration of privacy issues.

Pseudonymisation versus anonymisation

CBP, the Dutch DPA, has formulated a set of guidelines that need to be followed in order to offer pseudonymous services that are considered to prevent re-identification of individuals:

1. The first pseudonymisation takes place at the location of the primary data source.
2. Technical and organizational measures are adopted to prevent reproducibility of the pseudonymisation technique ('replay attack').
3. Processed data are not directly identifiable data.
4. An independent audit ex ante and at regular intervals shows conditions 1, 2 and 3 to be fulfilled.
5. The approach chosen is documented and made public.

The TTPs in health care all have implemented this procedure.
The most recent Opinion on anonymisation by the Article 29 Working Party (Opinion 05/2014) considers pseudonymisation to be a security measure, reducing the linkability of a dataset with the original identity of a data subject. It refers to anonymisation as coping with three major threats:

1. The threat to still single out an individual
2. The threat to link records relating to an individual
3. The threat to infer information concerning an individual.

It discusses a variety of anonymisation techniques, such as noise addition, permutation, differential privacy, k-anonymity, l-diversity and t-closeness. It concludes that it is possible to offer anonymity given present (combination of) techniques, but warns that the field is in flow, yielding new challenges over time.

5.2.5 Concluding remarks

The concise presentation of recent developments within the Dutch health care practices related to big data developments shows some interesting perspectives:

1. A trend towards commercialization of health care data is undeniable. The Dutch health care system by itself is a regulated market system in which regulated

competition has a place. Existing actors and new actors perceive additional opportunities to use health data for their core activities. All actors consulted are open to involvement of commercial parties such as the pharmaceutical industry.
2. Notwithstanding the interest in going more public and seeking for commercial exploitation of public data, all institutes clearly stick to serving the public interest of using health care data. Use of data for direct commercial purposes (such as marketing) is not supported. Requests from external parties that have these objectives in mind are not rewarded.
3. The internal use of personal data is very much focused on the core activities of the undertaking. This differs from data analysis to offering improved insurance services and forensics. Internal procedures are used that help preventing abuse and misuse of data.
4. All institutes that participated in this case study oblige parties to publish results of projects in which (de-personalised) medical data were used. This means that the result of the data processing activities should be of benefit for a larger audience.
5. All institutes within the case study use a TTP construction to deliver health care data to external parties. The TTPs (basically two: Custodix and ZorgTTP) use similar pseudonymisation techniques and procedures. The procedures are in line with the CBP guidelines for pseudonymisation. All institutes within the case study accept that this is not an optimal solution and that re-identification of individual persons can occur on an accidental basis. They do not have additional measures neither do they consider to implement additional measures to improve the pseudonymisation approach. They stick to the prescribed guidelines indicated by the CBP, being the legitimate Dutch DPA.
6. Some institutes think part of the solution will be in involving patients more directly in the loop (having control, providing informed and explicit consent). These institutes are open to alternative systems that enable these kinds of approaches. One party is reluctant given earlier experiences at other domains and the fact that this party is relatively distanced from the primary data sources. It expects additional complexities which may hamper innovation of health care.
7. Initiatives such as envisaged by the Almere Data Centre add a novel feature to the existing infrastructure on personal medical data. It adds a more commercially driven approach that tries to combine economies of scale with economies of scope, thereby hoping to contribute to effective and improved care. The initiative thus also copes with privacy issues that need to be solved in order to create a trustworthy environment for offering commercial services in dealing with personal medical data.

From a more general perspective we can make the following observations concerning the approach found in the case study related to our approach of a personal data market:

1. Personal medical data are not traded for purely financial gains by the institutes participating in this case study. On the contrary, one can notice reluctance to cross the line between using data with medical benefits in mind and a market approach in which (personal) data will be used for the benefit of one single commercial actor. The valuation of personal data is viewed from the perspective of quality of care/efficiency of the care system, and not from the perspective of monetary incentives.
2. Notwithstanding this non-monetary attitude with respect to the exploitation of personal data, institutes also realise that the business proposition of the past

decades – in which public funding assured the foundations of research activities aimed at health care improved – will have to be adapted in the light of changing governmental funding strategies and requires a more market oriented approach. Up till now, this market oriented approach is a prudent one.

3. With respect to safeguarding the privacy of patients, all institutes adhere to officially acknowledged and accepted safeguards (using TTP-constructs, fulfilling rights and obligations towards the patients as data subjects, using state of the art technology for pseudonymisation and anonymisation). Again, one can note reservations with respect to unduly use of personal data rather than seeking the edges of possible use within regulatory frames provided. This approach reflects fundamental rights to be higher valued than monetary value that is embedded in (depersonalised) medical data.
4. Finally, the involved actors realise that the discussion on the – commercial – use of personal data is not only related to the privacy of individuals but bear a more fundamental relation to issues of discrimination and profiling with it. Through use of data analytics, patient profiles may be developed that fire back on client populations and may have adverse consequences for access to health care services and re-imbursement of health care costs. This element of the introduction of data analytics in health care does not receive real attention today.

From the perspective of the model we developed in this study we can make the following observations:

1. We did not find many instantiations of the direct involvement of individuals in the data distribution model. The involvement of individuals was restricted to receiving bills (insurer) and to receiving some information on data provided to the researchers in the cohort studies (LifeLines).
2. In line with this first observation, the organisations in this case study acted as service provider towards third parties, rather than towards individual patients. In their relation with third parties they all dealt with shielding information of individual patients. The valorisation of data was clearly on the level of aggregated and pseudonymised data, and not on direct identifiable patient data.
3. The value of the data services that were provided, deal with non-monetary values, such as improved care and improved efficiency of care services. We did not find any valuation in terms of monetary value of patient data. On the contrary, much of the effort of the organisations in this case study is oriented at preventing direct monetisation of patient data.
4. We found some reference to the fundamental rights value of patient data. These relate to consequences of profiling and analytics that go beyond valuation for individual patients and are situated at a political level, related to non-discrimination and solidarity.

5.3 Services based on personal data in the financial sector

Organizations in the financial sector, such as banks and (payment) service providers, are developing data services based on client and payment data. Since this is a fairly new type of services for these organizations, the Nederlandse

Vereniging van Banken (NVB) recently published a position paper[29] on the use of aggregated client and transaction data by banks. While the NVB acknowledges the opportunities for data services, for example for fraud detection or personalized advertising, its core statement is that the privacy of clients always needs to remain secure and that clients always retain control over their personal data. In this section we will discuss the use of aggregated data for identifying patterns and anomalies and undertaking action based on these signals to determine a new code of conduct regarding the use of personal data, within the context of the Wet bescherming persoonsgegevens (Wbp, the Dutch Data protection act).
The NVB identified three categories of services in which client and payment transaction data can be used (according to the Wbp and the banks' code of conduct for the use of personal data[30]):

1. Creating client profiles for banking advertisements
   Payment transaction data, possibly contextualized with other data sources (postal code area, average payment amount, etc.), are used to create client profiles. Banks use these data on an individual basis for internal purposes. Based on these profiles, clients receive offerings of new banking products such as a higher interest rate or a new banking product.
2. Use of client profiles for offerings of external organizations
   Clients give their bank explicit consent to use their client and transaction or payment data for tailored relevant offerings of other companies. Data will not be sold to third parties; all data remain within the bank itself as the bank operates as an intermediary between the client and the third party offering a service or product.
3. Sharing client data with external organizations
   Based on the banks' code of conduct, two possibilities can be distinguished:
   - Market analyses based on transaction data. Based on data that are fully anonymized, reports are provided to third parties. As no personal data are involved, the Wbp is not applicable
   - Based on a specific ground, such as an obligation under the law, or fraud detection, or based on explicit consent by a client.

This use case investigates three services developed by organizations in the financial sector. Two services are developed by banks, while the third is developed by payment processing organization Equens. This organization, which is fully owned by banks, is the largest European payment processing organization for PIN and credit card transactions. Based on these payment data, the organization is exploring whether additional services can be developed. Their exploration is the first service that is investigated within this use case. The other services are developed by banks that also aim to develop services based on the payment data they store. The first is the MyOrder platform, owned by the Rabobank, and the second is the ING bank.[31] Two of the three services that are investigated in this study have gotten very bad publicity and have become controversial. The ING service has led to

[29] Nederlandse Vereniging van Banken, 'Position Paper Gebruik van Klantgegevens door Banken', 16 May 2014, www.nvb.nl.
[30] http://www.nvb.nl/publicaties/1691/gedragscode-verwerking-persoonsgegevens-financiele-instellingen.html
[31] http://www.ing.nl/nieuws/nieuws_en_persberichten/2014/03/ing_en_het_gebruik_van_klantdata.aspx?first_visit=true.

questions asked by Members of Parliament,[32] while the Equens service was nominated by the privacy rights organization Bits of Freedom for one of their *Big Brother awards.*[33] The negative publicity resulted in both projects being put on hold.[34] [35]

		Organisation-centric	Subject-centric
Added value incentive	Financial reward	Organisation uses data (including selling to third parties) and gives money in return	Data subject sells data for specific purposes in return for money
	Non-financial reward	Organisation uses data to add value, such as delivery or improvement of services and personalisation	Data subject provides data in return for services or personalisation
Fundamental rights incentive (control)		Organisation applies PETs / PbD to protect rights of data subject	Data subject controls his own data via the platform of the service provider

Figure 11: 'Services in the financial sector' category related to actor centricity and incentive

Figure 11 shows the mapping of these services onto the PDM classification provided in Figure 2. Considering all the attention this has brought upon these services, banks are likely to develop pilots in which individuals are in control over whether or not they will provide personal data for receiving these services.

5.3.1 Equens[36]

Equens was established in 2006 as a pan-European payment processor, through a merger of the Dutch payment processor Interpay and the German payment processor Transaktionsinstitut für Zahlungsverkehrsdienstleistungen. The three main Dutch banks (Rabobank, ABNAmro and ING) own 49% of Equens. The other 51% is under ownership of a German bank (DZ Bank, 31%) and an Italian bank (ICBPI, 20%), which joined Equens in 2008. In 2013, Equens processed 10,6 billion payments and 4,7 billion POS and ATM transactions.[37] On top of its core business, the processing of payments, Equens aims to develop additional services, such as mobile and biometric (based on fingerprints) payments. Furthermore, to capture the value of its payment data, the organization looked into the possibility of developing services for shop owners based on these data.

Equens' aim with their Equens Insights services was to compile better insight into customers of shops based on their payment data by creating aggregated profiles.[38] Shops can use these profiles to better target their customers, for example to create local market strategies, set up policies for locating shops and improving accessibility of shopping centers. Profiles will explicitly be created on an aggregated

[32] http://nos.nl/artikel/621368-kamer-wil-opheldering-ingplan.html
[33] https://www.bigbrotherawards.nl/2013/08/genomineerde-equens-zoekt-andere-manieren-om-de-rekening-te-betalen/.
[34] ING, 'Betreft: position paper ING t.b.v. ronde tafel "Gebruik klantgegevens door banken" dd. 21-5-2014, 14 May 2014.
[35] http://www.volkskrant.nl/vk/nl/2680/Economie/article/detail/3446572/2013/05/24/Equens-ziet-voorlopig-af-van-verkoop-pingegevens.dhtml.
[36] We are indebted to Dave Rietveld, head of innovation, business development of Equens for providing essential information on this case.
[37] http://www.equens.com/aboutus/companyprofile/key_figures.jsp.
[38] Equens Insights, 'Rapportage voorbeelden en inzicht in totstandkoming', 27 juni 2013.

level and not to generate information that can be traced back to individuals and their payment details, such as a link to the products they buy. Example of insights that can be created is the percentage of returning customers, the percentage of customers that generates a substantial amount of income, the percentage of customers that was 'lost' or 'gained', number of customers per weekday, the distribution of paying customers throughout the day, or other shops that are visited by customers of a specific shop.

The insights are generated by recording the payments that are made in shops, at the payment terminal (such as a PIN machine), the amount of the payment, the date and time of the payment, the type of shop (supermarket, or pharmacy), the payment pass reference (which is a unique anonymized encrypted number), and the distance between the location where the pass is registered (using the national CBS area code) and the payment terminal where the payment is made. Subsequently, the data is aggregated. In case too few (< 10) households are present in a specific CBS area, these are aggregated with other areas to ensure that data cannot be traced back to individuals. This method is comparable to the way of working of the CBS. Based on these aggregated data, graphs are created for the specific insights.

Equens maintains that this process was in accordance with the law, as the organization does not process personal data. (The organization does not store personal data and is therefore not held accountable by the CBP to the data protection law.) Equens compares their operations to the counting of traffic without registering specific number plates, as data streams flow unmonitored through their network. This means that aggregate numbers of cars (payments) can be registered without any reference to a specific car (payment). Hence, no data can be traced back to individuals. The organization compares itself to others services, such as the statistics bureau CBS that also 'counts' events, to the analysis of the payments that are made by the Rabobank, and search engines such as Google that also track views and data streams that pass through the nodes of their network. As the main reason for the service to be cancelled was bad publicity, Equens now aims to perform the analysis in cooperation with the CBS and consumer organizations to supply the data to those organizations rather than individual shop owners. By supplying them with free data, their reports can be improved, which also benefits Equens.

5.3.2 MyOrder (part of Rabobank)[39]

MyOrder, which can be seen as a service, an app, a platform, and a broker depending on which perspective is taken, used to be a startup company that was taken over by the Rabobank. The Rabobank, looking at a study performed by Accenture that stated that 32% of revenues of banks will disappear in the near future, concluded that a disruption in the world of finance will occur and that it needed to find new services to replace the decreasing revenues. The Rabobank – generally speaking – has three sources of income: payments, mortgages, and loans (mainly to SME's). MyOrder targets payments. Traditionally, payments are based on a 'four corner model' of guarantees of one bank to another that a merchant will

[39] We are indebted to Gertjan Rösken, CTO MyOrder, for providing essential information on this case.

get the money from a client. Banks earn fees in this model, for every transaction that is made. Such a fee is usually around € 0,02.

New payment services, such as Google Wallet, which is introduced in the US, are not based on the 'four corner model', but on capturing a percentage of a transaction that is completed. For example, a person has sought, using Google, for a specific type of jeans. When walking through a shopping street, the Google Wallet app in the smart phone that was used for this search, 'sees' that a shop in this street has these jeans on offer, and gives a signal, for example indicating that a discount can be gotten on these jeans. When the customer indeed buys these jeans using Google Wallet, Google gets a percentage of the transaction from the shop owner. With MyOrder, the Rabobank targets a similar business model. Rather than offering a mobile payment system that is only based on a transaction (such as the inclusion of NFC in mobile phones or payment cards), it is service based.

At this moment MyOrder offers a variety of services mainly targeting smaller and relatively simple transactions, such as the ordering of drinks at a café, the ordering of a bottle of champagne online, or payment of parking tickets. Payments can be made using different payment methods, such as credit cards, iDeal, or Minitix (a Rabobank specific online chipknip service). The app is not only for Rabobank customers, but anyone can download the app. Advantages for customers include not having to wait for a waiter to order a drink or for paying the bill and not having to buy a parking ticket for a fixed time (the remainder of the parking time – if any – is transferred back to the account of the user). As such it can be seen as a payment service or as an app in the app store or on a smart phone. At this moment 11.000 locations are linked to MyOrder.

Furthermore, MyOrder allows for cross-over services. One example is that multiple shops in a certain area could consider compensating visitors' parking costs using MyOrder, to attract more customers. Moreover, in the future MyOrder aims to become a platform for other services as well, allowing other service providers to develop services on the MyOrder platform. One example could be the integration of the MyOrder payment system with Albert Heijn's 'Appie' app that can be used to find products in a supermarket. The idea behind this is that a bank could well perform the role of a trust provider for payments and transactions.

The Rabobank will start developing user profiles based on the transaction data that it captures based on all the transactions using MyOrder. An example could be that one day a user gives a five star rating to a wine at a restaurant, and another day this, or a similar wine, is on sale at a local liquor store, and via MyOrder a signal can be sent to the potential consumer. In this way MyOrder, based on its user profiles, becomes a broker or intermediary. At this moment, users of MyOrder do receive messages about services that are available, such as parking services, but this is based on location data, not profiling. All these signals technically take place 'within the app'. According to legislation, banks are required to store transactions, once they have been performed. Thus, profiles based on these transactions can always be created.

In the current version, for registration at MyOrder, a mobile phone number is required. At registration, a Minitix account is opened without any money in it, but which may be necessary for reimbursing the user. When accessing MyOrder using

Facebook credentials, the service has access to personal data such as the name, address, and birthday. Often, also data such as the license plate number, and location data can be accessed and stored. Only data are captured that are relevant for a certain service. Personal data are not in any way sold, just used for profiling with the aim to offer profile-transaction based relevant products and services. This is the future business model for MyOrder. On a 'no cure, no pay'-basis, 'campaign managers' at other businesses can make offers of products and services via MyOrder. If a transaction is performed, MyOrder will take 5% of the value of the transaction up to a maximum of € 5.
At the moment, it is unclear what the market potential of these kind of services is. Google Wallet, for example, is not at all successful, but Google keeps running the service, because the company believes in this business model. Users will keep a certain degree of control over their data, as a 'cockpit' will be installed where users can turn on (and off) filters that allow business to make offers. As the Rabobank is required to store transaction data, these filters give only control to a certain extent, since they do not include deletion of transaction data. One interesting service where users also have control over their personal data is 'Sidekick', that is now tested in Leiden. This is a service within stores that allows tracking of customers through the stores. When a user 'checks in', he or she can receive all sorts of offers, and/or also receives a monetary reward (in the ballpark of € 0,05) for giving up personal data. Another option is that it will not be actual money, but a loyalty program. The underlying notion of MyOrder is that personal data belong to the consumer and they should have as much as possible control over their data. At least, this is now taking place in the form or choice.

MyOrder's ambition is not to have 16 million users, but to become a platform for transactions. The current setup is mainly a shop window showing the possibilities for other businesses that could build apps and services connecting to the platform. The Rabobank expects that this will be the real growth market. However, banks are quite behind in this world and need to catch up. On the other hand, banks could regain their trust function for guaranteeing payments. For a company like Google, this could be much harder to develop. At the same time, a threat to the banks is that organizations such as Post.nl and Vodafone are also developing platforms with similar functionalities. Compared to Vodafone, for example, banks may be much more conscious of prudent use of personal data as they have always performed a trust function. An internal bottleneck is the lack of funding for further developments.

5.3.3 Use case ING

Similar to the previous two cases, the ING bank aims to develop a service based on payment data that benefits their customers as well as generates income to companies aiming to improve their sales. In order to investigate whether this could be successful, the bank proposed a test involving a few thousand of their customers that would have had to give their explicit consent.[40] In case this pilot would have been successful, it would have scaled up to include all 4,2 million ING customers. While in the US these type of services are used by several banks, this is not the case in the Netherlands. As mentioned before, the service received very negative

[40] http://www.nrc.nl/nieuws/2014/03/11/wat-is-de-prijs-die-ing-betaalt-voor-het-delen-van-betaalgegevens/

reactions and the test was put on hold.[41] A graphical representation of the ING use case is presented in Figure 12.

Based on payment data as well as other data that is stored by the bank, ING creates profiles. Based on these profiles, third parties can specifically advertise to people with a specific profile. This would mean in practice, that the third parties supply the ING with these advertisements and the bank shows these profiles to their customers. Also, via the ING customers can purchase certain products, and the monetary gain is added to their account by the bank, not by the third party offering their product. Figure 12 shows that all personal data remain within the bank. This means that the third party is only in contact with the ING, and only gains insight into a customer data he or she purchases a certain product.

[41] ING, 'Betreft: position paper ING t.b.v. ronde tafel "Gebruik klantgegevens door banken" dd. 21-5-2014, 14 May 2014.

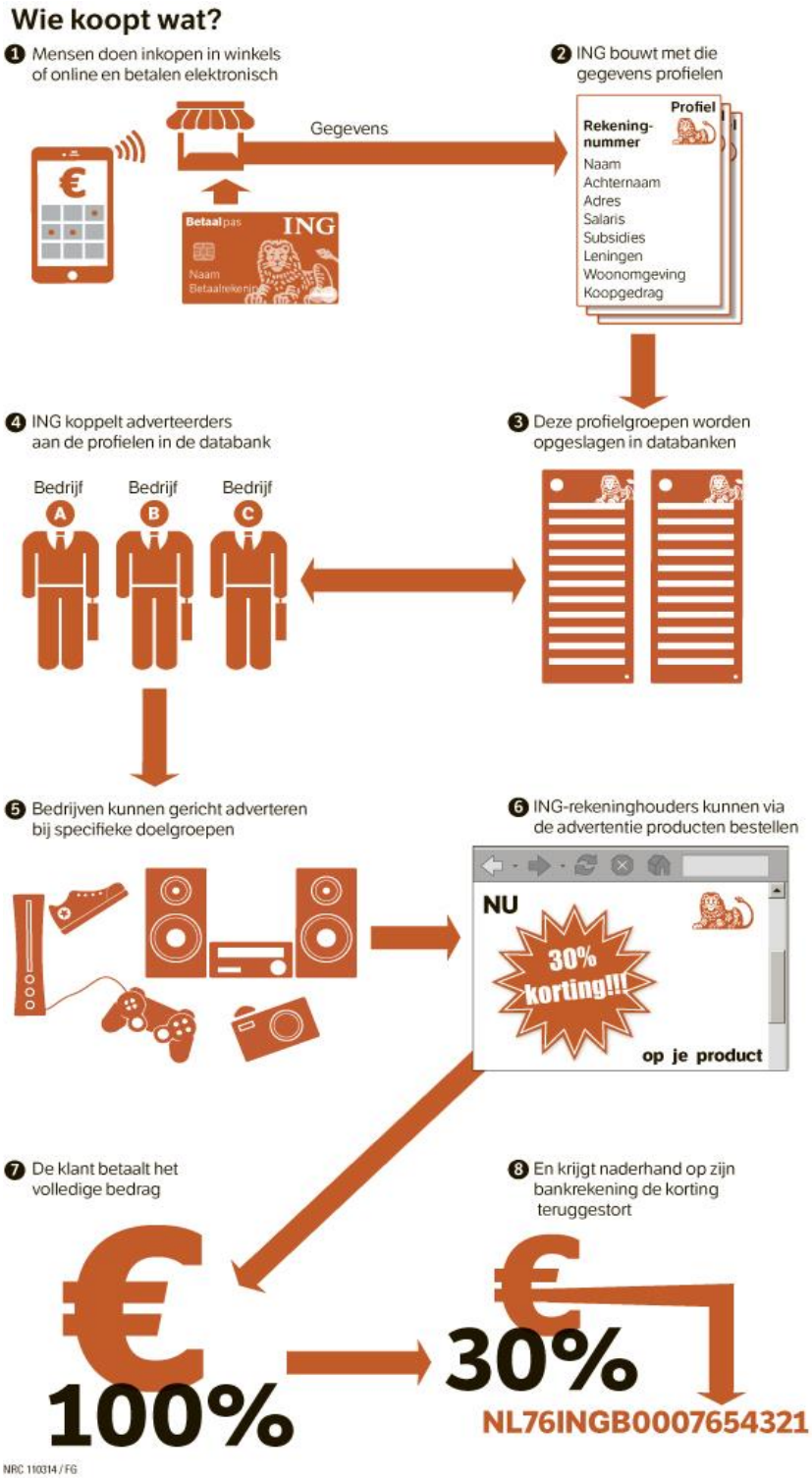

Figure 12: Schematic overview of the ING personal data service (Source: http://www.nrc.nl/nieuws/2014/03/11/wat-is-de-prijs-die-ing-betaalt-voor-het-delen-van-betaalgegevens/)

What is essential to the pilot project developed by ING that it only targets a group of customers that have indicated that they would like to participate in the trial project, via an explicit ‘opt-in’. This means that no personal data are used for this service of people that have not indicated that they would like to participate. A question that

arises is whether profiles can be created that are based on a limited number of people but that are still strong enough to generate a certain percentage of sales, which is something the advertising organizations are interested in. ING also uses big data for fraud detection, such as in case a bank card is skimmed and used two distant places at more or less the same time. For this, the bank uses the same data that are used to show the transactions to customers online or in apps.

5.3.4 Conclusion/reflection use case financials

Looking at the Code of Conduct published by the NVB, it appears that all three services are in line with these guidelines. However, as mentioned before, two out of the three services received bad publicity, which caused them to be put on hold. Secondly, what stands out in these use cases, is the similarity to services in other sectors, developed by other types of organizations. Vodafone, the phone company, for example provides similar services based on its client data, for example by providing traffic data using GPS data of mobile phones to satnav operator TomTom, but the company did not get bad publicity as a result. The question that arises is thus whether banks – as trust providers – should develop these services. The debate thus focuses on whether trust providers should enter data markets at all, considering their specific role. On the other hand, one also could state that whether banks are perhaps even better equipped to develop these services as they – more than other organizations that do not have such a role – are more used to the role as trust providers. This is a discussion that is likely to continue as organizations in the financial sector will explore their possibilities for developing data services.

5.4 Conclusion and reflection on use cases

As can be seen from the use cases and the forms of data usage presented in the previous sections, the upper half of our table is covered the most. The incentive of fundamental rights protection is almost absent. This may be related to the fact that there is no evident business model in the protection of personal data. One of the main reasons for this may be that the concrete value of privacy and personal data is not that tangible for individuals. As was seen in chapter two, the value individuals attribute to their data is very low as well. Nevertheless, we see that new innovative initiatives are on the rise, which take privacy protection as a service and develop systems to facilitate proper privacy protection, while enabling (new) services.

Despite the above, protection of personal data and privacy is a legal requirement. Organisations have to be compliant, so the challenge lies in developing privacy-respecting PDMs. This should allow for further development of services and new innovations, while at the same time protecting the fundamental rights and interests of individuals. A privacy-respecting approach towards PDM can be achieved wih the support of technological systems and frameworks. These will be discussed in the next chapter, in order to add a future perspective to the general overview provided in this report.

6 The role of privacy in the practice of PDM

6.1 The role of privacy

In the previous chapter we have presented models which are based on an added value perspective, either in the form of monetary incentives or in the form of improved or personalized services. Models based on a fundamental rights perspective are embedded in health practices. Outside the scope of the cases described is the approach offered by organisations such as Synergetics and QiY that offer radically different approaches to dealing with personal data, handing over control to the data subject while using an end to end trust assured framework (Synergetics) and organizing exchange of personal data in an advanced data locker system to which organisations can subscribe and on which data subjects remain in full control (Qiy). We did not take these approaches into account in this study, but for sake of completeness we mention these approaches as add-on to the ones we studied.

Providers of PDM need to cover their costs. A predominant business model in use is offering services for 'free' and using data of individuals for financial support of these services (selling data for advertisement purposes). Studies in behavioural economics have tried to 'calculate' the value of personal data in specific contexts for individuals (see chapter 2). This value shows to be very context-dependent, dependent on the kind of choices offered and on the precise configuration of the situation. As a baseline, willingness to pay for privacy protective measures is limited, while at the same time (potential) privacy infringements may lead to severe negative reactions of the public at large (see use case above).

Privacy is a fundamental right that needs to be respected. Because privacy is not straightforward embedded in PDM ecosystems, we propose some action in this area. Proper protection, basically, is necessary for PDM to be lawful. The embedding of privacy can be arranged by the use of technical implementations to protect data and to arrange control over the data. For instance, the use of a Trusted Third Party (TTP) or a trust framework can be of help in improving privacy protection. Both will be briefly elaborated upon below.

6.1.1 Trusted Third Parties

A commonly known solution for the provision of privacy is the use of a Trusted Third Party (TTP). A TTP is an organisation that fulfils a trust role between two parties. A TTP fulfils its role by adherence to technical and organisational instruments, such as a Public Key Infrastructure (PKI) and cryptographic key management. Within a PKI a Certification Authority (CA) authoritatively distributes certificates that are used to indicate trusted services. By using a combination of a publicly known and a private key, two parties can exchange any kind of message in order for these messages to only be decrypted by those parties in possession of the correct combination of the public and private key. This is a well-known and widely used approach for the trustworthy exchange of messages. Another instantiation of a TTP structure is a third party that fulfils pseudonymisation, storage and disposal of personal data when asked for under certain conditions and certain constraints. Within the domain of health care, TTPs are active that fulfil this role, usually by

means of hashing patient data. An approach often used is to strip personal data of identifying features (such as the zipcode) or to use more generic data instead of fully identifying ones (such as a person belonging to a specific age cohort instead of using the birthdate). The remaining data are hashed so that direct identification becomes more difficult.

TTPs can also be used to verify and authenticate persons. The TTP can hold a key which complements a key held by the individual data subject. The combination of the keys verifies the identity of the individual, without revealing the identity to the party asking for information. In relation to the verification of identity, anonymous credentials (Bichsel a.o., 2009) are a well-developed concept. The verification can also entail certain attributes or attribute based credentials, which indicate that the individual has a certain characteristic. For instance, the attribute that an individual is over 18 years old may be sufficient for certain transactions. The exact age of the individual can remain secret, since the only requirement may be that the individual is an adult. The TTP thus functions as an intermediary between two entities that both trust the TTP. The TTP knows the identity of both entities and confirms this towards both parties. By verifying the identity or attributes of the individual, a transaction with another entity is made possible without revealing the individual's identity, while both parties trust each other. A drawback of this system is that it involves an additional party into the transaction, next to the data subject, the service provider (as intermediary) and other third parties.

An example may highlight the function of a TTP. Since online behavioural advertising (OBA) is generally understood as privacy invasive, it may be helpful to include an anonymity system in the OBA advertising chain. An example of such an anonymity system is Adnostic (Toubiana a.o., 2010), which is designed as a browser extension and facilitates the targeting to take place in the browser of the individual. This means that the individual has more control in terms of opting out of OBA, and also that less data on browsing behaviour are communicated to other servers. Moreover, the ad-network does not know which ad was displayed to the individual.. As a result, the ad-network cannot derive information from the sequence of displayed ads either.

The system definitely contributes to privacy protection. However, there are also some drawbacks. Some of the drawbacks are rather technical, such as network latency and more intensive bandwidth use, click fraud, and the risk of less precise, and, thus, less commercially beneficial, targeting. One important issue, however, concerns billing. The fact that the final selection of an ad to be displayed is made in the browser of the individual, makes that the ad-network does not know which ad was displayed and, therefore, does not know which advertiser to bill. The authors of Adnostic propose a cryptographic billing system in which use is made of a TTP[42] which has to guarantee the anonymity of the individual. The TTP has to decrypt the ads to facilitate the correct billing. Two things become clear here: firstly, the TTP has a key position and has the ability (and task) to collect the targeted ad sequences, which emphasises the need for the trust aspect. And, secondly, the implementation of a privacy-friendly ad system has the consequence of the need for another system to facilitate correct billing. To construct a financially viable system,

[42] Or as the authors of Adnostic indicate, it is better to speak of a Trusted Sixth Party, since next to the user, four other parties are introduced before the TTP comes into play.

that is privacy friendly, one has to introduce additional parties that have a trusted relation to both the party offering a specific service (in this situation offering ads to people browsing the internet) and the individuals browsing the internet (who must rely that the TTP does not aggregate and sell the data collected on their browsing behaviour).

6.1.2 *Trust frameworks*

The relation between privacy (protection) and trust becomes very clear from the mentioned examples. Strikingly, trust is often sought by third parties which are not primarily involved in a transaction chain. It can also be seen that for commercial purposes, even in cases where privacy is protected by having analyses running on the individual's computer, at some point a connection must be made for billing purposes. Trust frameworks and TTPs are the most appropriate approaches to achieve this at the moment. Currently, several parties are working on the development of a solid and applicable trust framework. Examples of these are the Dutch Qiy Foundation and the Belgian SME Synergetics.

Below is a figure of a trust framework.

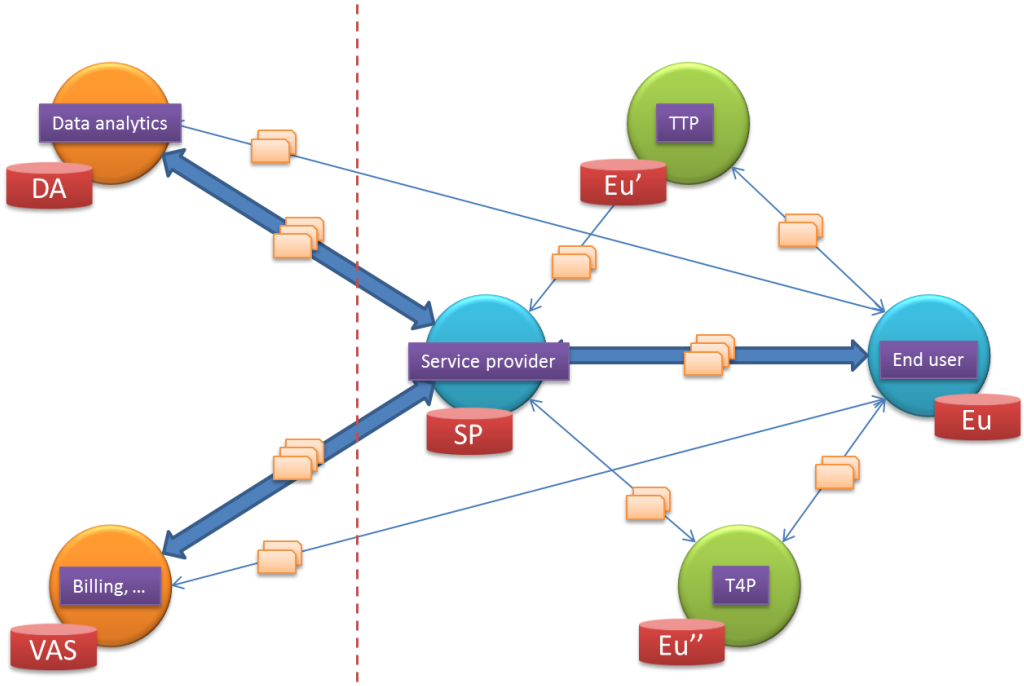

Figure 13: A PDM ecosystem as a trust framework

Explanation of the various components:

1. Key is the relation between the end-user (Eu) and the service provider (SP). The data exchange needed for the service provider to offer its services can be either directly organized between SP and end-user or indirectly through a TTP or a Trusted Fourth Party (TFP). Distinction between TTP and TFP is the closeness towards SP (TTP) or end user (TFP).

2. The SP has standing relations with external organisations that are either directly related to the service offered (improvement of service, billing, ...) or indirectly through data analytics. The kind of data flowing between SP and data analytics organization/billing organization will depend upon the service delivered by these actors. They are themselves embedded in a network with other actors as well, adding to the complexity of the PDM ecosystem.
3. SP: Service Provider
   DA: Data analytics
   VAS: Value Added Services
   Eu: End user
   TTP: Trusted Third Party
   TFP: Trusted Fourth Party
   Thick blue line: Main data flows
   Thin blue line: Secondary data flows.

The market value of the personal data as this can be derived from this figure thus depends upon the role of TTP and TFP and the terms of agreement between end user and service provider.

7 Conclusions

In this report we started by addressing the emergence of personal data markets. Forecasting reports show that PDMs are on the rise, and that their economic impact is considerable. Figures of 8% of GDP in Europe are mentioned, a figure that should already be reached in 2020 (BCG 2012). A large fraction of this market, so do US market reports show, relates to the market of personalised advertisement, or Individual Level Consumer Data (ILCD). This market segment is already 50% of total advertisement market volume in the US (Deighton and Johnson 2013). So, personal data markets are already a reality, and may fit the scheme the World Economic Forum has drafted to differentiate between various categories of data: personally provided, observed and inferred data (WEF 2014). Though our study did not detail the origins of the ILCD, other studies show that a large portion of US based marketing data may stem from observed data and even personally provided data without clear consent and choice of the data subjects.[43] In this study we attempted to determine the value of personal data from the perspective of the firm (value of the firm at the stock market, valuation of revenues per data record, costs of data breaches per record) and from the perspective of the data subject (behavioural economics). The economic approach helps understanding part of the business equation to be made by a single firm, while the behavioural economic approaches help understanding some basic and core psychological features of how individuals value their personal data. All in all, research in this field has limited prospective value, and no hard claims can be made to how individuals value their personal data and the precise conditions that they consider to be relevant in safeguarding or negotiating these personal data.

The practice of PDM is still limited. Some examples exist, but there is no substantial market yet. Nevertheless, it seems that step by step new initiatives see the light and larger companies as well as small start-ups try to find a proper way to establish a PDM. The (potential) value of personal data is big and recognized. This is an incentive for PDM. At the same time, PDM is seen as a chance to provide privacy protection. A more ideological viewpoint is at the basis of these initiatives, but the business model is less evident then as well. The challenge is to establish a PDM service which protects privacy and at the same time facilitates new services.

Some of the examples presented in this study indicate public resilience – or may be one should say negative media attention – against untethered use of personal data. This sometimes led to the rejection of initiatives that – on closer examination – do take basic protective features into account and could be considered as an attempt to include privacy constraints in the business equation. So, even if an initiative for a PDM practice is fully compliant with data protection requirements, public perception can be decisive in whether the initiative can be successful or not.

In attempting to overcome the constraints of evaluating media attention as such, this study presented a framework that may help evaluating which features of personal data in a personal data market are of relevance for a typical situation. We differentiated between organisation centric and user (data subject) centric

[43] See for instance ‘Apps en privacy’, *Computeridee*,no. 8, Volume 2011.

approaches, and concluded on the basis of several examples that the organisation centric approach is most dominant. The data subject hardly plays a role in the business process that leads to offering a service to this data subject. Usually, the data subject is a passive recipient of the service, while the negotiation on his or her personal data takes place in other arenas. We differentiated between two main perspectives: an added value perspective, where the added value takes the form of a specific monetary value or a non-monetary value (improvement of service delivery for instance), and a fundamental rights perspective (in which personal data are valued on their relationship with fundamental rights).

In the examples we studied we found all three perspectives present. We presented one example of a subject oriented approach in which the data subject himself negotiated the – monetary – value of his data. This, however, is far from a typical situation, and it was very much oriented towards raising awareness for the mere fact that personal data represent monetary value. The other two main perspectives found were the data subject oriented one who delivers data in return for services (financial sector) and the organisation centric one in which the organisation uses data of data subjects for non-monetary incentives as well. This is not to say that monetary incentives do not play a role. The advertisement industry which we briefly discussed is a clear example of an industry that runs on the monetary valuation of personal data, usually without the direct interference of the data subject.

The organisational structure of Personal Data Markets which we presented, is composed of three main components: the data subject, the service provider and a third party, external to the direct relation between service provider and data subject. On the basis of the analysis of the various cases, it showed that the third party can be helpful in the organisation of the information exchange between the service provider and the data subject, for instance by offering trusted services that help protecting and safeguarding personal data of the data subject. Dependent on the position of this party it is either labelled a trusted third party (being closer to the service provider) or a trusted fourth party (being closer to the data subject). Other third parties that do not bear the label ‘trusted’ can offer a variety of services, usually for the service provider, such as billing and data analytics.

With respect to privacy protective measures, several options are available. The most far-reaching is the end-to-end trusted framework between data subject and service provider. This is hardly found in common practice. It is under development in niche markets by niche actors that are creating a full ecosystem that should help data subjects to achieve privacy (in terms of transparency, control and choice) while it helps service providers in creating a trusted infrastructure for offering their services (including demonstrating accountability). Other solutions, especially by means of trusted third parties are much more common. We found the reliance on TTPs in the health care sector as part of normal practice, following guidelines as offered by the Dutch DPA.

The use of these TTPs has, however, not led to widespread use and commercialisation of, in this case, patient data. Even with these safeguards in place, one can notice reluctance to use patient data outside the realm of non-monetary valuation, i.e. of improving the quality and efficiency of healthcare. Cross-overs to the monetary valuation of patient data are avoided by the health care actors that participated in this case study, Patient awareness, fear for negative

publicity and a reluctance to provide data for the benefit of (other) commercial actors are part of the motives provided.

In the other case study on financial services we noticed the impact of negative publicity on services within a PDM that nevertheless were compliant with Dutch legislation on the use of personal data. Negative media exposure led to halting two initiatives that tried to reconcile commercialisation of personal data with data protection approaches.

The case studies thus revealed some interesting additional aspects to our study of PDMs:

1. PDMs in settings where financial motives are predominant, flourish but are largely invisible to the public. These PDMs function on collection and aggregation of personal data, using all three data categories (personally provided, observed and inferred). The Online Behavioural Advertising market no doubt is the largest segment of these markets.
2. PDMs in settings where non-financial motives are pre-dominant, sometimes face larger media exposure than the first category, notwithstanding privacy protective and respecting measures taken by the service providers in these PDMs.
3. Privacy preserving features that can be introduced in PDMs to safeguard privacy interests are available, are already in place (TTPs) or are 'under construction' by niche players that orient themselves to partial or integral approaches of safeguarding the relation between service provider and data subject.

From a broader societal perspective we arrive at the following conclusions:

1. The development of PDMs to provide new or improved services that may have beneficial impacts for society at large (such as improved health care) is sometimes hampered, because of the fear for negative side-effects (in commercial or in media terms). This appears to be particularly the case when sensitive data (health data, financial data) are processed in the PDM.
2. At the same time, warnings are phrased against societally disturbing consequences of large scale use of personal data.

Where the first conclusion opens the door to looking for innovation practices that reconcile privacy awareness with societally beneficial uses of personal data, the second conclusion refers to the need for a more in-depth political deliberation on the longer term potential impacts of the creation of (privacy-preserving) PDMs.

8 References

ACQUISTI, A. 2009. Nudging privacy: The behavioural economics of personal information. IEEE Security and Privacy, 72-75.

ACQUISTI, A. & GROSSKLAGS, J. 2007. What can behavioral economics teach us about privacy?, in: A. Acquisti, S. Gritzalis, S. Di Vimercati, C. Lambrinoudakis (Eds.), "Digital Privacy: Theory, Technologies, and Practices," Auerbach Publications, pp. 363-379. ISBN: 1420052179.

BCG 2012. The value of our digital identity. Boston Consulting Group.

BICHSEL, P., CAMENISH, J., GROSS, T., SHOUP, V. 2009. Anonymous Credential on a Standard Java Card. In: Proceedings of the 16th ACM conference on Computer and communications security, pp. 600-610. New York: ACM.

DEIGHTON, J. & JOHNSON, P. A. 2013. The Value of Data: Consequences for insight, innovation and efficiency in the US economy. Harvard.

ENISA 2012. Study on monetizing privacy. An economic model for pricing online information. Brussels: ENISA.

FINN, RachelR. L., David Wright, and MichaelD., Friedewald, M. 2013. 'Seven types of privacy', in Serge Gutwirth, Yves Poullet et al. (eds.), *European Data Protection: Coming of Age*, Springer, Dordrecht, 2013.

GUTWIRTH S, GEHLERT R., BELLANOVA R. et al. 2011. D1. Legal, social, economic and ethical conceptualisations of privacy and data protection, FP&7FP7-PRESCIENT.

HILDEBRANDT, M., O'HARA, K. & WAIDNER, M. 2013. The value of personal data, Amsterdam, Berlin, Tokyo, Washington, IOS Press.

KATIBLOO, F. 2011. Personal Identity Management.

NEUTEBOOM, O.B.E. & SIJMONS, J.G. 2012. Rapport Big data in de zorg: geheimhouding en privacy, Molengraaf instituut Utrecht.

OECD 2013. Exploring the Economics of Personal Data - A survey of methodologies for measuring monetary value. In: OECD (ed.) OECD Digital Economy Papers.

RABIN, M. 2013. Incorporating limited rationality into economics. The Journal of Economic Literature, 51, 528-43.

REGAN P.M., Legislating Privacy: Technology, Social Values, and Public Policy, University of North Carolina Press, Chapel Hill, 1995, p. 221

SOLOVE D.J. (2008. Understanding privacy. Cambridge, mass., London, England, Harvard UP.

SPIEKERMANN, S. 2012. Privacy property and personal information markets. Acatech - Deutsche Academie der Wissenschaften. Berlin.

THALER, R. 1980. Towards a positive theory of consumer choice. Journal of Economic Behavior and Organization, 1, 39-60.

TNS-NIPO 2011. Attitudes on Data Protection and Electronic Identity in the European Union. In: EUROBAROMETER (ed.). Brussels: European Commission.

TOUBIANA, V., NARAYANAN, A., BONEH, D., NISSENBAUM, H., BAROCAS, S., 2010. Adnostic: Privacy Preserving Targeted Advertising. Appeared at NDSS 2010. Online available at: http://crypto.stanford.edu/adnostic/adnostic-ndss.pdf.

VEEN, E.B. van 2012. Centrale opslag van ‘Big Data’ zonder toestemming van de patient mag niet (zonder meer). MedLaw Consult.

WESTIN, A. 1967. Privacy and Freedom,. New York,: Atheneum.

WORLD ECONOMIC FORUM 2010. Personal Data: the emergence of a new asset class.

WORLD ECONOMIC FORUM 2013. Unlocking the value of personal data: from collection to usage. World Economic Forum.

ZARSKY, T. & ANDRADE, N. de 2013. Regulating Electronic Identity Intermediaries: The 'Soft eID' Conundrum. Ohio State Law Journal, Vol. 74, No. 6, 2013.