IPACSO White Paper

The Marketplace for Privacy-related Products and Services

by Dr. Nicola Jentzsch

March 2015

Abstract

With the large-scale deployment of Information and Communication Technologies markets for products and services related to personal data and privacy are thriving. Yet, to date, there has been little effort to systematically map these markets. In this IPACSO White Paper, a methodology is proposed that enables a systematic horizontal as well as vertical analysis of markets for personal data and privacy-related products and services. This method is applied to classify some of the major players active in these markets.

IPACSO White Paper No. 1 (March 2015)

This White Paper is produced as part of the Innovation Framework for Privacy and Cyber Security Market Opportunities (IPACSO). It draws partially on the material in the Reports D4.1 and D4.2A. It does not reflect the opinion of DIW Berlin as institution or IPACSO as consortium, but the author's own.

Corresponding Author

Dr. Nicola Jentzsch, DIW Berlin, Mohrenstr. 58, 10117 Berlin, T. +49 30 89789-234, F. +49 30 89789-103, njentzsch@diw.de

Table of Contents

List of Figures

List of Tables

I. Introduction

In the past 30 years advances in ICT deployment have led to a large-scale increase in the collection and processing of personal data.[1] This has led to thriving markets for products that consist of/or are based upon personal data and their analysis. Examples include direct marketing and credit reporting, where millions of consumer profiles are sold on a daily basis. The wide range of players that compile user profiles from different sources in order to sell them for different purposes is expanding. Moreover, services that yield as by-product personal data are also expanding.

Personal data products and services serve the goal of establishing or maintaining control rights over personal information in different contexts in order to increase personal privacy. Examples include delisting and Internet reputation management services (e.g. unroll.me, godelete.com, reputation.com), as well as personal data vaults (such as personal.com).

Other products promise privacy by explicitly refraining from the collection of personal data. Examples are anonymization websites including TOR and PirateBrowser, but also Guerrilla mail and Enigmail.[2] For those companies anonymization is a competitive differentiator. However, there are also products that are versioned and that privacy-sensitive customers can adjust to their preferences (for example websites that allow a change of privacy settings).

At this stage we have no systematic categorization of service providers, the size of the markets, market segments, or how competition in these markets work. One reason for this lack could be that market identification and sizing is a difficult task: Privacy, i.e., either the absence of personal data or the confidentiality and integrity of personal data, is polymorphous. It can emerge as a transaction quality or as a product trait (see Figure 1 in chapter 2). This makes it difficult to judge whether a service provider or product/service is part of the

[1] Personal data in this context is understood as defined by *EU's Data Protection Directive* 95/46/EC.

[2] Further detail is provided in D2.3 (Section 1.2)

market. After all, privacy settings are now also integrated in Facebook and Google, without being independent, stand-alone products.

Three major trends are observable in this context. The first is that an increasing number of products and services can be traced back to individuals, for instance, consider the advancement of electronic payment means versus cash. Digital finance is an expansion of the market of identity-related products and services in finance.

The second trend is that even large corporations, which are part of the World Economic Forum, are now starting to rethink the traditional way of building customer relations. This culminates in the statement that the customer has to be put at the center of data control, because individuals are becoming more savvy and sophisticated about data processing practices.[3]

The third trend is a new type of start-up company or privacy innovator providing technologies for data reduction and/or data control. Examples are companies such as Blackphone (SGP Technologies) and Sedicii. The UK consultancy Crtl-Shift estimates that there are about 400 firms internationally active in the market for such products and services.

One major concern raised by the absence of market mapping and sizing is that we cannot monitor and track the developments in these markets. As of February 2015, there is only anecdotal evidence concerning market developments. On the other hand, there is a vast (academic) literature on legal and economic aspects regarding personal privacy. At this stage, we do not know who the main players in these markets are, what products are provided, or how the market is segmented. We can also no estimate how the market is evolving.

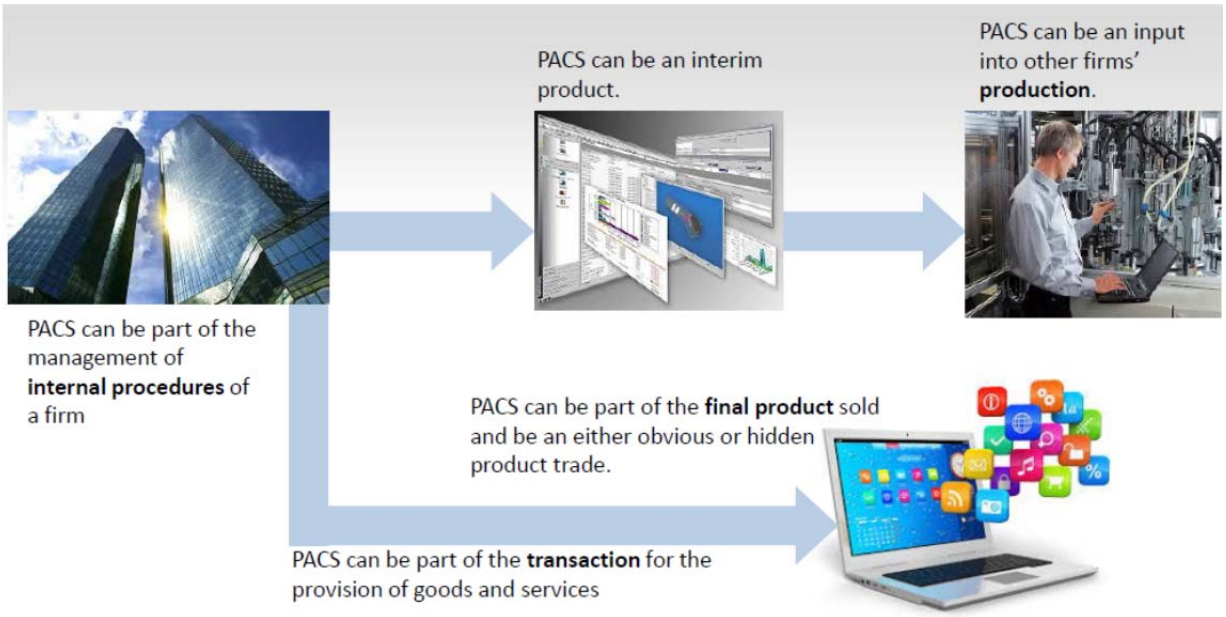
This White Paper is only a first step toward the development of a more comprehensive taxonomy of markets, where demand and supply for personal data and privacy products and services meet. A robust taxonomy is a potentially useful tool for market analyses conducted by firms, researchers, and regulatory authorities. Such a tool would not only allow to size the market, compare markets across countries and to follow trends, but also to analyze the market's interaction with other markets, such as the market for ICT products and services.

[3] See World Economic Forum website: http://www.weforum.org/projects/rethinking-personal-data

II. Identification as Product Input

A major challenge faced when trying to map the markets for personal data products and services is that privacy is polymorphous.[4] It can, as stated in the introduction, surface as business practice, a transaction quality, or a product trait (see Figure 1).

Figure 1 Privacy and Cyber-security (PACS) in Economic Transactions



Source: Jentzsch (2015).

For example, there are stand-alone products (such as the Blackphone), which are clearly part of the market of products that provide increased privacy. However, there are also privacy features integrated in Apple's iPhone, which have to be turned on for privacy-enhancement. This complexity makes it difficult to judge whether a service provider or product is part of the market or not. After all, privacy settings are now also integrated in many products without being independent, stand-alone products.

[4] For simplicity we will use the term product for both products and services.

Before we discuss ways of market identification and segmentation, one important aspect must be stated at the outset. It seems that official statistical sources cannot be used for the identification and assessment of the market. In Europe, there is the standard industrial activity classification NACE (*Nomenclature statistique des activités économiques dans la Communauté européenne*). This classification system does not capture privacy-related software and services, because this is not the level of precision provided in the industry classification system.

A first step is to look what classifications have been used in adjacent markets; for example the market for cyber-security products. Note that identification of this market is as difficult, but there have been more trials. A frequent separation made in cyber-security studies is to separate the market into hardware, software and services (for an overview over these studies, see Jentzsch 2015a). This segmentation can be associated with relevant classifications in the NACE system.

Table 1 Market Segments in Cyber-security and Privacy Markets

Markets	Horizontal Market Segments		
	Hardware	Software	Services
Cyber-security	X	X	X
Privacy		X	X

Source: Jentzsch (2015a).

According to Jentzsch (2015a), markets for personal data and privacy products can be preliminarily separated into a software and a services segment (see Table 1), but there seems to be no hardware production segment. Instead, hardware is used as production input in markets for privacy-related products, where either more secure products are used[5] or common-off-the-shelf technology is used.

[5] An example is hardware computers with higher emission security.

2.1 MARKET DEFINITION

The market for privacy-related products can be defined as a physical or virtual place, where the supply and demand for such goods and services meet.

Relevant players in these markets are companies whose primary activity is related to the provision of software tools or services related to the protection of the digital identity of individuals. "Primary activity" is the activity of the company that generates *most of its revenues* by either providing anonymization or pseudonymisation tools or by providing products based upon data collection and analysis in order to increase data control or to reduce personal data disclosure.

Note that the *primary purpose* of the product must be a*n increase of personal information control in order to enhance privacy*. Products that are based upon personal data collection and analysis and that do not have this explicit goal do not belong to the market. Note that there are many products that promise greater information control, examples are personal analytics tools, including Wolfram|Alpha Personal Analytics for Facebook, JawBone, Apple Health, and Digifit.[6] However, it is the combination of both, control and privacy that yields products that would belong to the market as defined here.

Thus, we can state that the market for privacy-related products is the physical or virtual place, where supply and demand for goods and services meet that primarily serve the purpose of personal information control in order to increase privacy.

[6] An overview of such tools is provided here: http://quantifiedself.com/guide/

2.2 IDENTIFICATION DEGREE AS PRODUCT DIFFERENTIATION

One possibility to segment the market for privacy-related products is the degree of identification needed for the production of the product. Why the degree of identification? The act of identification is a key competitive differentiator for persons who strive to maintain or increase their privacy.

Privacy has become a key selling point and many companies use and some misuse the term. While some firms need personal identification as production input, others do not. Thus, an anonymized version and a personalized version increase the options of choice for the consumer. Moreover, these different business models put firms under completely different regulatory regimes: whereas firms that collect personal data need to adhere to data protection laws, firms that do not collect personal data do not fall under such regimes. And, of course, there is a grey area, where it is unclear whether data are personal or not (IP numbers, for instance).[7]

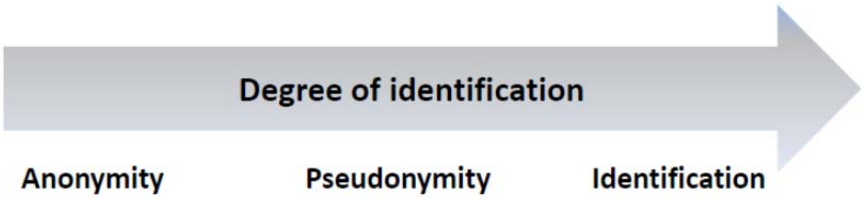
Consider service providers that collect identification data in order to provide products, which are related to increased information control. Examples include personal data vaults and delisting services. For these, identification *is necessary* in order to provide the service of storing information or for delisting a person from different websites.

Other companies provide anonymization services. Examples include TOR or StartPage. Here, no identification is needed, because the providers do not collect personal data or computer IP numbers as quasi-identifiers.

Because identification is so important, we should more fundamentally look at it. Identification should not be regarded as a binomial (yes/no, 0-1) variable. There is a whole range of different degrees of identification (see Figure 2).

[7] Google disputes that IP addresses can always be seen as personal data (see Google blog entry at http://googlepublicpolicy.blogspot.de/2008/02/are-ip-addresses-personal.html)

Figure 2 Range of Identification Degrees



Source: The author.

A greater level of resolution allows us to differentiate the products from each other and to identify different market segments. At the same time, it allows us to draw conclusions on what products could be substitutes once viewed from the user's perspective. Note that the above scheme is a proposal, other methodologies of segmentation are possible as well and the identification degree is just one element in the consideration of whether products are substitutes or not.

The interesting side aspect is that there is a mathematical concept of anonymity: The degree of identification increases with the probability of being drawn from an anonymous set of subjects (Diaz et al. 2002). This enables an understanding of "what kind of anonymity" a product provides and whether one type of anonymity is a substitute for another.

Some of the products (with and without identification of the person using it) might serve the same purpose (veiling the identity of a subject), although they use different inputs for achieving this goal. Again, as soon as persons are identified with their natural or legal identity, a different regulatory regime is applicable. A firm that does *not* collect personal data does *not* fall under data protection laws.

Thus, we can identify the following market segments, which are often lumped together under the header "privacy market:"

(1) Anonymity products:[8] At the one extreme end, there is the segment of anonymity products, where the suppliers do not require any personal data input from their customers;

[8] Many IT-experts agree that there is no guarantee of 100% anonymity if digital technologies are used. Thus, it is more apt to speak about 'quasi-anonymity.' However, for simplicity the world anonymity is used in this report.

(2) Pseudonymity products: In the middle range, we find products that allow customers to use fictitious names (pseudonymity) and quasi-identifiers (IP numbers);[9] and

(3) Identity products: Products that require personal data input either directly by the customer or indirectly.

While the re-identification risk posed by anonymity products is minimized (it is not zero, though), it is 1 or almost 1 for identity products. In the following, this segmentation will be used for the description of two different supply chains in order to make the point more clear.

Consider the two generic supply chains models that are differentiated by the degree of identification used as production input. In Figure 3 (A), the production input is the disclosure of personal data by the individual. Note again that this is simplified, because there are other production inputs that are ignored here, for expository purposes. One example is Google web search based upon a user's IP number.

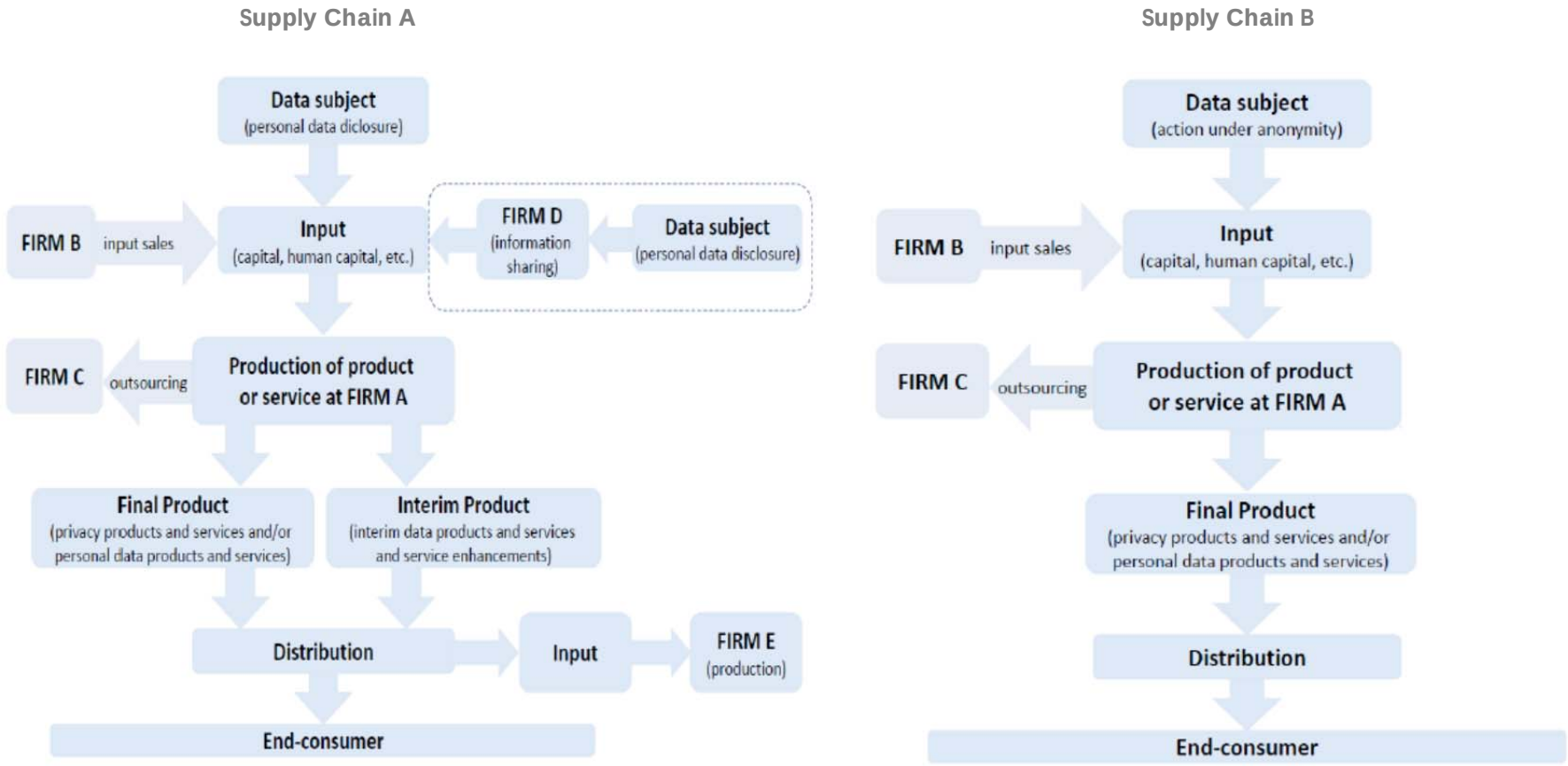
Supply chain B, by contrast, does not require the act of identification.[10] Examples are the StartPage and Ixquick search machines, which do not record IP numbers. Although 100% anonymity cannot be guaranteed, these services make re-identification harder.

Note that products can be also through-put, i.e. they serve as production inputs into the provision of other products. An example is credit reports for the provision of credits by banks. The market segment of interim products is very large, if you just consider the credit reporting or micro-marketing business.

[9] MySpace used to be an example of this approach – persons could log in under false names.

[10] Note that identification is not always an active and conscious act of the data subject, but might be indirect identification by third parties.

Figure 3 Supply Chain in Personal Data and Privacy Markets



Source: Jentzsch (2015b)

2.3 Other Market Segmentation Methods

Other segmentation methods are thinkable as well. Consider the intuitive segmentation along different types of data (health data, credit data, marketing data, etc.). Such a method has been applied by some national competition authorities in Europe.[11,12,13] While this is very intuitive, the dynamics in markets for personal data, however, lead to a convergence of some segments and to an alteration of others. Therefore, market segmentations need to adapt to the dynamically changing substitution relationships between data products. For example, the analysis of social network data allows inferences of creditworthiness, although they are not traditional credit repayment data.[14] Thus, in future such data and combinations thereof might be used as substitutes for traditional credit repayment data.

Another method might be to differentiate on the type of data collection, i.e., whether data have been volunteered, observed, or inferred (see OECD 2013). The consultancy company Crtl-shift (2014: 31) focuses on information products, where consumers possess controlling rights. The three types of segments identified in this market are:

(1) Personal Data Management;

(2) Decision Support; and

(3) Life Management.[15]

The first type helps individuals to gather, store and analyze their own data (e.g. personal data vaults). The second enables individuals to collect and use information to make better individual decisions (such as price-comparison

[11] Office of Fair Trading (2004). Completed Acquisition of Acxiom Corporation of Clarita Europe Group, including Claritas (UK) Ltd.

[12] Office of Fair Trading (2004). Anticipated acquisition by Acxiom Corporation of Consodata SA.

[13] Bundeskartellamt (2005). Beschlussabteilung B9 – 32/05 (2005). http://www.bundeskartellamt.de/SharedDocs/Entscheidung/DE/Entscheidungen/Fusionskontrolle/2005/B9-32-05.html

[14] See for an example U.S. patent US20140129420 (http://www.google.com/patents/US20140129420).

[15] An overview is provided in the infographic at: https://www.ctrl-shift.co.uk/news/2014/07/30/pims-infographic/

machines) and the third type enables individuals to use information to manage life tasks and processes (such as moving homes).

In fact, a combination of the above approaches could yield the most precise results for market identification and mapping. Consider the differentiation along the degree of identification, type of data collection and served purpose of the product. For the latter, it is important whether the product is seen as a substitute in the eyes of the user. It can be assumed that a person requiring an anonymity product will not regard a product as substitute, if he/she needs to provide personal data in order to obtain it.

If the turning-on of the privacy features of a product renders it inter-changeable with another product, both can be seen as substitutes and, thus, rival offers.

All in all, seen from a macro-perspective, anonymity products increase customer choice and therefore potentially improve on consumer welfare.

III. Classification of Business Models

It is at this stage not possible to conduct a full-scale market analysis. This must be left for future research. Alone the identification of information sources from which knowledge on market players could be drawn is a study in itself. What can be provided is an initial classification of market players using the proposed methodology. As stated, the quality of identification of the customer differs across players and business models.

Table 2 classifies the businesses along four main dimensions: (1) the identification of the customer, (2) the monetary compensation rules in the transaction, (3) the non-monetary compensation rules; (4) and third-party data sharing. Note that while all of the firms listed in the Table are active players in the privacy-related product marketplace, firms such as Google, Groupon and BlueKai earn the main share of revenues from services other than the ones related to privacy. They are still listed in order to differentiate them from other business models.

Identification and Identification Proxies

As discussed above, business models vary with respect to the quality of identification they need in order to work. While some only require proxy identifiers, others revert to the natural identity of the data subject. Technical proxy identifiers include, for example, IP addresses, mobile phone IMEI information, numbered bank accounts, and user names. They shield the "natural identity" of a person to a certain extent, although not perfectly. Legal identity includes real name, address, date of birth, as well as identifiers such as ID card numbers. This information allows individuals to be uniquely identified, at least with the power that a state has to identify its citizens.

Table 2 Business Model Matrix

Business Model Matrix	Product Qualities		Pecuniary incentives		Non-pecuniary incentives
Company	Identification		"Positive" Prices	"Negative" Prices	Personalization
	Identification Proxies	Legal Identity	Earnings	Payments	
Google	X				X
FaceBook	X				X
Experian		X		X	X
TransUnion		X		X	X
Equifax		X		X	X
Acxiom	(X)				X
Foursquare	(X)				X
Reputation				X	
Handshake			X		

Note: In credit reporting (Experian, TransUnion and Equifax), banks typically price in the price of a consumer report requested on the applicant. While a good risk gets lower prices, a high risk will get a mark-up for credit taken up. The net effect depends on the circumstances.

(X) denotes that once these firms also collect ID numbers (or SSN in the U.S.), they would have the legal identification of data subjects. The author has no information whether the firms in fact collect this information.

Monetary Incentives and other Incentives

The second important distinction when looking at business models is how individuals are incentivized to disclose personal data. Are they directly being offered a monetary benefit (e.g. the Handshake business model) or do they obtain reduced prices in form of discounts (e.g. Groupon business model)?

Do they need to *pay* in order to "control" their information better (e.g. Reputation.com, credit reporting)? These compensation mechanisms are quite different to non-monetary social exchange, where non-monetary incentives play a role. If data are disclosed for some other benefit such as a personalized search engine, website or other service (e.g. Google, Twitter, Facebook), non-monetary incentives play a role in the exchange, such as reciprocity and fairness.

Third-party Data Sharing

It is noticeable that in the most successful business models, the data subject does not actively take part in the secondary transaction, where their data are commodified and monetized. Examples of these models are online advertisement companies (Google, Facebook) as well as credit reporting agencies (Equifax, Experian, TransUnion) and direct marketing firms (Acxiom). A reason could be that users are often uninformed about the terms-of-trade of their data and the money earned from the data is not obvious to the data subject. Moreover, data subjects could have an incentive to strategically modify their information or not to disclose it once they would have a say in the transaction (an example is credit reporting).

There seems to be an increasing number of start-ups that provide consumer-direct services, where consumers directly participate in the revenue sharing from third-party data sales. These companies are also called data vaults, data lockers or personal information management systems.

Firms may offer one specific tool or service, or, in some cases, they are active in several of the above segments by offering a portfolio or products and services. For example, the US- and UK-based credit-reporting agency Experian

offers credit and marketing reports as well as self-monitoring products. On the other hand, Ixquick, an anonymous search engine is only active in that very segment.

IV. Conclusions

This White Paper proposes a methodology to identify markets for privacy-related products and services. Such market identification is important in order to enable market sizing and trend monitoring. A market is a place, where demand and supply for goods and services meet. In the case of privacy-related products it is the market for products that serve the primary purpose of personal data control to increase privacy. Firms that obtain a majority share of their revenues from selling such products and services are considered to be active players in this market. As stated, however, this is just a first step in the direction of better identifying these markets.

It was stated that one possible criteria of segmenting the market is the degree of identification of the customer/user needed for the provision of the product or service, which has the goal to increase privacy. The premise is that customers choose the product based upon their privacy preferences. Additional criteria, such as type of data collected (health, financial, etc.) and type of data collection (direct, indirect), will increase the precision of market segmentation, though.

This approach here is used to classify firms and their business models. Not all of the classified firms are active players in the privacy-related product marketplace. But this exercise shows that the companies can be classified along the proposed method.

Future research must focus on further detailing the above and it should be devoted to the identification of sources from which company information can be drawn and statistics can be compiled. Without this information, policymakers cannot track these markets and their development.

List of References

Jentzsch, N. (2015a). State-of-the-art of the Economics of Cyber-security and Privacy, IPACSO Report Deliverable D4.1, published on ipacso.eu

Jentzsch, N. (2015b). Horizontal and Vertical Analysis of Privacy and Cyber-Security Markets, IPACSO Report Deliverable D4.2 A, published on ipacso.eu

Diaz, C. S. Stefaan, J. Claessens, B. Preneel (2002). Toward measuring anonymity, Presentation at the PET'02, www.csl.mtu.edu/cs6461/www/Slide/Measuring%20Anonymity02.pdf