Exploring the economic value of personal information from firms' financial statements

Claudio Feijóo , José Luis Gómez-Barroso , Peter Voigt

A B S T R A C T

Currently personal data gathering in online markets is done on a far larger scale and much cheaper and faster than ever before. Within this scenario, a number of highly relevant companies for whom personal data is the key factor of production have emerged. However, up to now, the corresponding economic analysis has been restricted primarily to a qualitative perspective linked to privacy issues. Precisely, this paper seeks to shed light on the quantitative perspective, approximating the value of personal information for those companies that base their business model on this new type of asset. In the absence of any systematic research or methodology on the subject, an ad hoc procedure is developed in this paper. It starts with the examination of the accounts of a number of key players in online markets. This inspection first aims to determine whether the value of personal information databases is somehow reflected in the firms' books, and second to define performance measures able to capture this value. After discussing the strengths and weaknesses of possible approaches, the method that performs best under several criteria (revenue per data record) is selected. From here, an estimation of the net present value of personal data is derived, as well as a slight digression into regional differences in the economic value of personal information.

1. Introduction

Since marketing techniques initially made the importance of client-focused strategies clear, companies have been collecting customer data and using them to create value. When there was *just a real world*, those data were difficult to collect and were stored inside the firm "as the miser watches over his hoarded gold" (Douplitzky, 2009). The Internet, or rather information and communication technologies (ICTs) have radically changed this situation. Currently, data gathering is done on a far larger scale and much cheaper and faster than ever before. While we communicate, exchange and access information using electronic communications systems, data records can be and are collected on who we are, where we are, what we do, and how we do it. Even more significant, personal data[1] are easily shared or transferred across companies, markets and industries, irrespective of geographical boundaries.

Arguably, the most striking feature of this scenario is the emergence of a particular type of company for whom personal data are the key factor of production. The business models of quite a number of companies operating in online markets – including many of the Internet giants – are based on targeted advertising, which in turn relies on behavioural profiling. Personal data are thus becoming one of the main assets of many modern markets, to the point that they can be considered "the new oil of the internet and the new currency of the digital world" (Kuneva, 2009).

[1] In this paper the terms "personal data" and "personal information" are used synonymously. For both, we assume the definition of the term according to the OECD Privacy Guidelines: "Any information relating to an identified or identifiable individual (data subject)". Typical examples of personal data include, for instance, name and address information, social security, health or other unique identifying numbers, health records and financial information. However, these days, techniques can often enable data relating to search terms, websites visited, GPS positions, and IP addresses, to be linked back to an identifiable individual. Thus, the paper assumes a broad understanding of personal information, but does not attempt a clear delineation from non-personal information.

Surprisingly, this type of statement is still based largely on theoretical grounds. Few would deny that personal information generates value for companies but, at the same time, almost no academic paper can be produced to help estimate how large this value is. Until now, the economic analysis of personal information has been conducted from a qualitative perspective, mainly linked to privacy issues (a review of the economic literature on information privacy can be found in Pavlou, 2011). However, there is a pressing need to progress from those qualitative statements to quantitative results. This is not least because some of the main data collectors are traded on stock markets and are quickly ascending the list of the biggest world companies.

This article seeks to shed light on this topic, trying to approximate the value of personal information for companies, in particular for those companies that make their business precisely through the handling of personal information. In the absence of any systematic research or methodology on the subject of study, an ad hoc procedure has been developed. The accounts of a number of key players in online markets have been examined. This inspection has tried first to determine whether the value of personal information databases is somehow reflected in firms' books, and second to define performance measures able to capture – alternatively or complementarily – this value. After discussing the strengths and weaknesses of these possible approaches to valuing personal data, both from a conceptual and practical (obtaining the figures for key players) perspective, the method that performs best is selected. Taking those values as a departure point, and after further processing, an estimation of the value of personal data is derived.

Along these lines, the article is structured as follows. After this brief introduction, the next section introduces the complexities involved in the valuation of personal data. The use of company reporting for the valuation of personal data, both from a theoretical and practical perspective, is analyzed in detail in the following section. From this discussion, the results obtained through the different possible methods employed are displayed, as well as discussed in terms of their different merits. Finally, the paper closes with the calculation on the net present value of personal information, a brief investigation of regional differences and some conclusions on the relevance and applicability of the valuations obtained in the paper.

2. Background – the difficulties to value personal data

Regardless of the purpose for which it has been collected, personal data that a firm holds are one of the intangible assets it administers. Therefore, at a conceptual level and in principle, the problems arising when trying to give personal data a value are the same as for any other intangible assets: mainly that the requirements for reporting them are few and often imprecise (Cohen, 2005; Hunter, Webster, & Wyatt, 2012) and, when existing, place limits on the managements' ability to record them,[2] given the uncertainty associated with payoffs from the assets and information asymmetry surrounding managements incentives (Wyatt, 2005; Grosu et al., 2011). "Even worse, much of the information that is provided is partial, inconsistent, and confusing, leading to significant costs to companies, to investors, and to society as a whole" (Lev, 2003).

Such problems and concerns have frequently led analysts to look at stock markets as an independent valuation source that provides "seemingly objective assessments of a firm's intangible assets" (Whitwell, Lukas, & Hill, 2007). Intangible assets are one of the possible contributors to the disparity between market and book values of companies (Hulten & Hao, 2008). Therefore, market capitalization should somehow reflect the value of intangible assets. Following this reasoning, researchers and stock analysts have poured their efforts into evaluating how intangible capital shows up in the valuation of the firm without agreeing on the preeminence of any specific methodology. Such a task is not easy to achieve. Gu and Wang (2005) state that high information complexity of intangible assets increases the difficulty for analysts to assimilate information and they find that forecasting errors are greater for firms with diverse and innovative technologies. Market bubbles and volatility can also lead to mistakes or dubious outcomes (Bond & Cummins, 2000).

Among the various intangible assets, the link between R&D investments and market value has been a major focus of research and there are many studies following this avenue for research, from Griliches (1981) to Hall and Oriani (2006), Ehie and Olibe (2010) or Palmon and Yezegel (2012). In terms of customer value, the number of studies that have addressed the topic is much smaller (Berger et al., 2006; Gupta & Zeithaml, 2006; Kumar & Shah, 2009; Schulze, Skiera, & Wiesel, 2012). All these articles study how customer equity (the total combined customer lifetime values of all of a firm's customers) contributes to market value. Note that customer equity is derived from the role of clients as present and future buyers of products and/or services from the firm. To the best of our knowledge, only one previous and very recent work has valued clients in their role of "personal data donors". Cauwels and Sornette (2012) derive a valuation of Facebook and Groupon from the evolution of three scenarios for their customer base. Apart from this work, there are no studies that value personal data as an independent category of intangible assets and it can be asserted that the work of valuing personal data is by and large still to be done.

From both an academic and economic perspective, this is not an acceptable situation as a growing number of companies rely on business models that are largely dependent on the use of personal information (Gómez-Barroso & Feijóo, 2013). For those companies, personal records databases can – should – become visible in book accounts. Acknowledging the problems exposed at the beginning of this section, this is unquestionably a first immediate way to look at the value of personal data. The "market way" is also valid: regardless of decisions made by accountants, the market approach can be used in the same manner as it is used for any intangible asset. The next section explores both paths – not necessarily confronted[3] – presenting a number of options for approximating the monetary value of personal data. As will be discussed later, each of the options presented potentially suffers from certain methodological biases and measurement errors. The approximations obtained are therefore not directly comparable; that is, they may complement each other rather than being substitutes. Yet, put in context and seen altogether, they can still serve as an initial basis for expanding the overall understanding of the economics of personal data.

3. Method – valuation of personal information based on company reporting

Below, six different approaches to deriving a valuation of personal information are analyzed. The analysis is undertaken with regard to those companies that ground their business entirely or to a major extent on the commercial use of personal data. In particular, the empirical part focuses in detail on five companies: Experian, Facebook, Google, LinkedIn, and Xing, although the preferred method is later extended to five additional companies in Appendix

[2] For instance, goodwill is permitted by most accounting standards to be recognized as an asset when it is acquired as part of another business, but this is not the case when it is internally generated.

[3] Choi et al. (2000) state that the financial market positively values reported intangible assets.

II. Facebook is the largest social network in the world. LinkedIn and Xing are networks for professionals. Experian is "the leading global information services company, providing data and analytical tools to clients around the world". Google is the dominant search engine and increasingly a provider of several "cloud-oriented" services and applications.

The first set of methods looks at the value of data stock indicated in the companies' financial reports, either directly (as such) or indirectly (through the analysis of "other intangible assets" or goodwill entries). The second group – the last three methods in the analysis below – approximates the value of data by means of relating company performance indicators (market capitalization, revenue, net income) to the stock of data held.

All these approaches will focus on the value assigned to an individual record or user. Each of these records corresponds to "a profile". A profile could contain as little as a single piece of information for an individual, such as the person's age, address, etc., but more typically represents an entire set with combinations of demographic, economic and educational information, and thus contains much more comprehensive information. Companies build such profiles from the information gathered in a single transaction or in the course of multiple transactions, potentially adding other pieces of information collected – or deduced – outside the specific context of the transaction.

We should note here that, apart from the particular limitations for each method described below, a general limitation affects all equally: the amount of personal data reported as being held by a certain company might be subject to fluctuations and, moreover, eventually also to "strategic reporting".[4]

3.1. A notional approach

3.1.1. Stock of personal data reported as company assets

As stated in the previous section, companies that rely entirely/mainly on the commercial use of personal information may want to capitalize the stock of data, i.e. report them as company assets. In such a case, monetization of an individual profile record would be straightforward as it could be obtained directly from the company reports and/or the balance sheet.

Even if this were so, it would not be a completely reliable method. First, concerns about strategic reporting apply. Companies may not publicly disclose the real value they assign to the data records they hold by as this may lead to biased/misleading signals concerning the retail price of data affecting the firms' position. It is a matter of fact that a public valuation of the companies' data would be critical for the reporting company since, on the one hand, the assumed overall value of data capitalized as company assets would significantly change their balance sheets[5] and, on the other hand, eventually provide sensitive information about their core business to clients, competitors and public bodies.[6]

Second, the price to be paid for gaining access to certain data significantly depends on the context of use of these data, the sector, the rival use of the data, etc. In particular, in the case of non-rivalry in terms of use, a certain record might be used or sold several times without affecting the revenues from each individual commercialization. More complicated cases could arise with partial non-rivalry, where the value of the data decreases with each commercialization. Hence, a comprehensive approximation of the full value of a certain data record needs to take this into account and discount all future revenues potentially to be made with this data record to the present.

3.1.2. Intangible assets per data record

Even if the value of the data is not explicitly reported, it might still be captured by "(other) intangible assets" of the company, assuming that databases of personal records are a fundamental part of the unspecified assets. In that case, by relating the balance sheet position, presumably containing the value of the database (e.g. "other intangible assets") to the stock of data, an indicator of the individual economic value of a data record might be obtained. Crucial in this context is to sort out for each company's annual reports and financial statements what exactly each individual position comprises (i.e. double-checking the footnotes and the corresponding auditors' comments). The latter is important since, as already noted, the reporting of intangibles across companies currently is not undertaken in a coherent way; moreover, it might be subject to strategic company decisions. In other words, due to the standards currently applied and the common praxis in terms of company accounting and reporting, this method may provide a biased picture.

3.1.3. Goodwill per data record

Since firms' goodwill[7] represents the difference between market capitalization or the market value of a company and its assets, the resulting measure would represent an approximation of the non-tangible/not reported value of the company. The latter, in the case of firms grounding their business entirely or overwhelmingly on the use of personal information, presumably corresponds widely to the economic value of these data. Yet, the drawback is that a number of variables which have little or nothing to do with the value of the data records – such as brand image or expectations – may affect a company's market capitalization and therefore also the reported figures of the firm's goodwill. In addition, a transaction in which goodwill appears would be required to make it apparent.

3.1.4. Market capitalization per data record

Assuming sufficient information on the firm is available, market capitalization reflects the value of the firm with all future earnings considered and discounted to their present value. This is the reason why it can differ – even significantly – from book value, which is a not forward-looking snapshot. Therefore, in theory this could be a valid approach. However, in practice, the market value of a firm often fluctuates with the general market sentiment and/or according to other economic shocks that may or may not be tied to the underlying value of the personal data. In addition, the firm should be on some stock market to allow for the availability of reliable audited information.

3.1.5. Revenue per data record

Revenue per data record is a means of capturing the "productivity" of an individual record. The approximation may be considered a more realistic measure of the monetary value of data than that

[4] Some "degrees of freedom" in this regard seem to lie in companies' approach to the number of valuable profiles they hold, for instance, in the reporting of "inactive accounts" as "current users" (i.e. profiles not actively accessed for more than six months, but not deleted). The metric related to "monthly active users, MAU" is becoming increasingly popular as a means of defining what an active user is.

[5] This would imply a balance sheet extension, i.e. virtually inflating the reported equity/company assets while the income would remain unchanged with the effect of nominally decreasing the "return on equity" ratio (ROE). As this ratio is a key financial indicator, company reporting which it affects is seen as sensitive for companies seeking new investors and amplified access to finance.

[6] In fact, the pure indication of what kind of data and in what detail is available to the company (which in turn may allow the sources of the data and the means of processing, etc., to be assumed) may already go beyond the amount of information the company wants to disclose vis-à-vis competitors, clients and public bodies.

[7] Goodwill is the extra a potential buyer of the firm would be willing to pay for those assets not appearing in the balance sheet, such as, for instance, knowledge capital, intangibles such as customer contacts, but also elements such as brand, image, reputation, etc.

based on market capitalization because it is less likely to be affected by changes in stock markets. However, revenue per data record illustrates the average return for a certain data record in a given reporting period (usually, one year); i.e. looking at it ex-post rather than capturing ex-ante the possible future revenues and earnings, corresponding to the concept of pricing on the stock markets. In fact, the latter needs to be taken into account and discounted to the present value in order to obtain a comprehensive approximation of the "true" economic value of the data record. As stated before, doing this is not trivial as it implies anticipating any potential future commercialization of the data record (implicitly prices, profit margins, etc.) and also the period of amortization (i.e. when the data will become outdated or obsolete). Again, publicly available and reliable information is required.

3.1.6. Net income per data record

This figure represents revenues minus costs, normalized per data record held by the firm. Costs may thus include expenditures for acquiring additional data or accessing certain information the company does not hold itself but needs to complement its own data resources. In this light, (net)profits per record – due to their nature of being an aggregate of revenues and costs – might be a good indicator of the "profitability" of personal information records held by a firm. Against this, the argument presented above concerning the need to consider and discount any future revenue holds also for profit per record. Moreover, assumptions made in terms of costs (data mining, processing, maintenance, updating) become crucial here. Arguably, this will be highly context-dependent; i.e. in relation to the sector and also probably firm-specific. For instance, social networks, while relying on personal data provided and kept up to date "voluntarily" by the account holders, will have different costs in terms of data maintenance as a data broker, doing its own data mining, merging and matching of several data sources. Complete visibility of costs is also more difficult even in the case of firms listed on stock markets.

3.2. An empirical approach

For illustration, the above-mentioned measures are compared for some shortlisted companies (Experian, Facebook, Google, LinkedIn, and Xing AG) in order to evaluate the plausibility and robustness of the corresponding approximations (see Appendix I).

It was indeed difficult to identify further candidates to be added to the sample. Actually, companies needed should mainly – or ideally exclusively – generate their revenues from data services and, moreover, disclose their economic figures e.g. by issuing comprehensive annual reports or the like. The smaller firms usually do not do the latter. Other sources of data can be used but, without prejudicing other possible approaches for further research, companies shortlisted on stock markets provide information that is more reliable, coherent and comparable than any other publicly available source.[8] As for the first observation, other firms may obtain obviously revenues from data services too but do not disentangle this for accounting/reporting purposes from further business domains which they may have running in parallel. In other words, aggregation level of information is a further challenge. In any case, five additional companies have been identified as potential candidates for the application of this empirical approach. A short description and the results from the preferred method are presented in Appendix II.

3.2.1. Personal data reported as company assets

For the time being, companies remain reluctant to provide details concerning the amount of data they hold and even more so with regard to communicating estimates of the corresponding economic value. Only in one case, Experian, a value associated to the database held is reported among the company assets.

In fact, Experian, in its Annual Report for 2012, displays at position (21) "other intangible assets", the sub-category "databases". The company states that "capitalized databases comprise the fair value of databases acquired as part of a business combination or the data purchase and data capture costs of internally developed databases. Databases are held at cost and are amortized on a straight line basis over three to seven years". In other words, the databases are not capitalized at their "true" economic value (including discounted anticipated/future revenues) but at the cost of data purchase/data mining, processing/maintaining and updating, and considering the depreciation over the mentioned three to seven year period.

The company's databases are reported to have a net book value of US$458 million (on 31 March 2012), compared to US$428 million and US$393 million in the previous business years, respectively. The estimates of this position are fairly stable and remain widely unaffected by market shocks. Considering the number of data records, the value of each personal datum would result in a value of US$0.69 and US$0.71 for the years 2012 and 2011, respectively.

3.2.2. Intangible assets per data record

Among the companies considered, Experian is the only case to report explicitly relevant intangible assets related to databases of personal information, as discussed above. Google and Xing display intangible assets in their accounting, leading to valuations per data record relatively congruent to those of Experian (see Appendix I for figures). However, LinkedIn has a considerably lower value of intangible assets per data record, US$0.06 for the year 2011, hinting at intangible assets as too vague a category[9] to be used congruently across firms for a stable comparison. Even more so, Facebook, as discussed below, blends intangible assets and goodwill in the same accounting category, further complicating the use of this approach for practical measures to undertake an economic valuation of personal information.

3.2.3. Goodwill per data record

The figures in Appendix I suggest that the goodwill of some firms is fluctuating, akin to the figures of market capitalization. See, for instance, the goodwill per record for Facebook in Q1 and Q2 in 2011 compared to the same time period in 2012 (US$0.19 vs. US$0.85), which arguably cannot be driven by the (rather smooth) increase in the number of user accounts.[10] The same holds for Xing.

Again, only Experian seems to have rather stable figures over time; these appear to be surprisingly close to the figures of revenue per data record. The latter, in turn, indicates that the firm's goodwill cannot include a "fair valuation" of the firm's own databases – if any at all – since this would imply anticipating future earnings due to the data and discounting these to their present value.

3.2.4. Market capitalization per data record

Appendix I shows that market capitalization per user profile is subject to considerable fluctuations. The case of Facebook is

[8] Note that Google does not report data on number of users. Therefore, an estimation was prepared on the basis of penetration of the Internet, penetration of search engines and market share of Google across regions. Official statistics were used where possible and complemented with industry analysts' data were needed.

[9] Intangible assets may also include, for instance, human capital and thus particularly accumulated (tacit) knowledge, designs/corporate ID/value of company image, the economic value of established access to markets/distribution channels (relevant, for example, in the case of access barriers/sunk costs), etc.

[10] The number of monthly active users rose in total from 845 to 955 (+13%) (http://allfacebook.de/userdata).

particularly instructive. In February 2012, Facebook prepared an initial public offering (IPO) of its stock which put the valuation of the company (market capitalization) at about US$104 billion (US$38 per share). This works out to a valuation equivalent to US$112 per registered user for a user base of roughly 900 million. On the day of the IPO, 18 May 2012, this figure rose to even more than US$120 per account. Moreover, the market capitalization – and thus implicitly market capitalization per record too – kept fluctuating quite significantly, as has also been the case since the IPO. Just three months after the IPO, in August 2012, a Facebook user account would be valued at about US$55–US$60.

At the end of 2011, LinkedIn showed a market capitalization per user account of approximately US$43. In the case of Xing, it was about US$25. The figures appear to be of similar magnitude and may indicate a corresponding confidence interval for the valuation of the corresponding personal data.

The figure is also similar for Experian, which showed at that time a capitalization of around US$23 per record. Experian has been listed on the stock market since October 2006, and this can help test whether the company figures are driven by internal values or by market sentiment: Experian's market capitalization per record fluctuated in a pattern that is similar to overall market trends between 2007 and 2011.[11]

3.2.5. Revenue per data record

Although Facebook, LinkedIn, and Xing, on the one hand, and Experian on the other, employ very different business models, and store different types of data, which they also collect/obtain in a different way, it is interesting that these firms reported fairly similar levels of revenue per user or record. In fact, at the end of 2011, the figures calculated for the mentioned companies' revenues per user/data record appear surprisingly close at about US$4.60 for Facebook (Q1 and Q2 in 2011 extrapolated), US$3.60 for LinkedIn, US$7.40 for Xing and US$6.80 for Experian. These figures point towards some robustness of valuating data by means of revenues. Data in Appendix II for an additional set of companies contribute to this picture of relatively coherent valuations.

3.2.6. Net income per data record

Two pieces of evidence support the fact that net income per data record arguably does not reflect the economic value of the data accounts: the net income per data record for LinkedIn between 2007 and 2011 was found to be fluctuating around zero (for some years even becoming negative); in the same line, the net income per record for Facebook as reported for Q1 and Q2 in 2012 (financial statement released 26 July 2012) dropped to almost zero (US$0.05).

4. Discussion

As for the theoretical approach, without yet having had a look at the numbers, revenues per record arguably seem to be the most robust indicator for approximating the value of data from the company's financial information. This indicator captures how much the firm earned (in total) due to data activities, while market capitalization and profit per record would include costs and/or external factors such as market sentiment that could fluctuate and, moreover, may not be connected to the underlying value of the data (ceteris paribus and assuming that the firm concerned is mainly/entirely doing business on the basis of their stock of personal data). As for the alternatives that directly use entries of intangible assets, they do not seem to be superior as they can suffer a number of potential biases including hard dependence on the decisions taken – in particular strategic reporting – and the absence of a forward-looking perspective.

In fact, the evidence presented indicates that market capitalization and goodwill fluctuate significantly and, as predicted, these measures do not appear to be driven primarily by the value of the personal information but by a number of factors such as market sentiment. Evidence also suggests that approximating data value by means of profit per record tends to be misleading. The same holds for intangibles if the value of the database is capitalized at cost. In sum, in terms of indicating the economic value of personal data, revenue per record seems to be better suited than the other options, or is at least the most robust option.

It should be underlined once again that when using this indicator it is assumed that firms' revenues come predominantly or exclusively from data-driven business activities. This is the case for the five companies included in the sample. Apparently, Google is the one that could have broader revenue sources but during the years of the study, according to the Annual Reports no less of 96% of revenues come from advertising (relevant, cost-effective online advertising as Google defines it), i.e. from personal data, and this was considered in the calculations. As discussed later in the final section, the relationship between the volume of personal data, and the financial results of a company may suffer from several drawbacks. However, the rest of indicators are also exposed to the same problems and, therefore, a more complex schema combining various indicators would not make results any better.

5. Further calculation – towards the real value of personal information

5.1. Net present value of personal profiles

In terms of the previous discussion, revenue per record seems to provide a better gauge of the monetary value of personal data. Yet revenue per data record remains a performance figure (implicitly a productivity measure) which still needs to be turned into a reliable economic value by also considering future returns.

Accordingly, a multiplier needs to be derived starting from the present value of revenue per data record. This calculation requires a number of assumptions. The simplest case supposes the permanence of the firms' fundamental business concept, lack of ad hoc shocks, no need to acquire additional data to complement the existing data, no current costs for maintaining the commercially usable data, non-rivalry in the usage of profiles, and constant usefulness of profiles during the time span for the calculations below. Considering all this and applying the aggregated Net Present Value method [NPV], the formula to be used is:

$$R_{NPV} = \sum_{t=0}^{T} \frac{R_{to} - (t/T)R_{to}}{(1+i)^t}$$

where R_{to} is the current revenue per record, i is the interest rate, and T is the maximum lifetime of the data record.

T, the time frame in which existing data become economically obsolete (depreciation horizon), is crucial in this formula. There are no data in the literature about the lifetime of personal profiles in terms of economic value. Only from the consumers' perspective a recent study in the UK about personal information on e-commerce and search transactions showed that users do differentiate their valuation when storage of personal information shifts from 2 to 5 years but not beyond (Potoglou, Patil, Gijón-Tascón, Palacios, & Feijóo, 2013). However, this hint is not operational enough and

[11] Looking at valuations per record/user (approximated by market cap) for Experian and Facebook, normalized with the trend of the Dow Jones Industrial Average index (DJIA), revealed that both companies follow a similar path but trail the overall recovery shown in the US by the DJIA (i.e. underperform in terms of valuations per record/user compared to the overall market trend).

Table 1
Influence of interest rate in profile valuation.

Assumed reference benchmark interest rate	Revenue per record	0%		5%		7%		10%	
		Data value min	Data value max	Data value min	Data value max	Data value min	Data value max	Data value min	Data value max
Facebook	4.3	8.54	17.08	8.27	15.55	8.17	15.02	8.03	14.30
LinkedIn	3.6	7.20	14.40	6.97	13.11	6.89	12.66	6.77	12.06
Xing	7.4	14.80	29.60	14.34	26.94	14.17	26.03	13.92	24.79
Experian	6.8	13.60	27.20	13.17	24.76	13.02	23.92	12.79	22.78
Google	21.7	43.30	86.61	41.94	78.83	41.45	76.16	40.74	72.53

Table 2
Regional differences in net present value of profiles (in US$, 2011).

Region	Mean revenue per record	US & Canada	Europe	Asia	Rest of the world
Facebook[a]	4.3	9.5	4.9	1.8	1.4
LinkedIn	3.6	6.0[b]	–	–	2.0
Xing	7.4	–	17.5[c]	–	2.0
Experian	6.8	10.7[d]	14.8[e]	–	3.60
Google	21.7	77.3[f]	28.4	28.3[g]	3.6

[a] For Facebook growth statistics/penetration per world region, see http://www.internetworldstats.com/facebook.htm.
[b] US only.
[c] Germany only. 2011 German share approximated from 2010 as it appears in: http://corporate.xing.com/fileadmin/user_upload/XING_AG_US_IR_Presentation_US_Roadshow_April2011.pdf.
[d] Revenues from the US and Canada, but data records from US only. Reference for data records: http://www.experian.com/small-business/sales-leads.jsp.
[e] Revenues from the UK and Ireland, but data records from the UK only. Reference for data records: http://www.experian.co.uk/marketing-services/consumer-data-and-insight.html.
[f] US only.
[g] Only the main members of APAC: Australia, China, Hong Kong, India, Japan, Malaysia, New Zealand, Singapore, South Korea, Taiwan, and Thailand.

therefore, it has been preferred to use data from the industry. As discussed, Experian assumes that the firm's databases are amortized on a straight line basis over three to seven years. This would mean than after three to seven years, the data from personal profiles lose their usefulness to extract economic value, that is, to provide personalized services through the specific business model of the firm. Thus, those values are taken. The influence of the interest rate i is summarized in Table 1. For instance, if the benchmark interests were assumed to be at 5% or 10%, the multipliers would be 1.94 and 3.64 (5%) or 1.88 and 3.35 (10%) of the current revenue for the upper and lower bounds of the confidence interval (three or seven years' amortization assumed), respectively.

Note that converting these values into total revenues is subject to the uncertainties in the future evolution in the number of users. According to Cauwels and Sornette (2012) in their study on social networks, a clear sign of a change in regime occurred in 2010 in terms of the growth in the number of users, from pure exponential behaviour (a paradigm for unlimited growth) to a logistic function with asymptotic plateau (a paradigm for growth in competition with some market saturation or maturity stage). Following their estimations, which consider revenues of US$3.5 per user per year, even in the case of extreme growth in the number of users of Facebook, revenues per user should increase by a factor of 1.5–3 in order to meet the current widespread high expectations of stock markets.

5.2. A digression into regional differences

It is worth mentioning that relating any company performance figure (such as revenue) or measures (such as goodwill) to the total (global) amount of accounts only makes sense if every account is assumed to be worth the same economic value. However, there is reason to believe that there are considerable regional differences. For instance, data concerning a European account holder compared to one from the US, Asia or elsewhere might commercially be of a different value; that is, in an advertising business model, the retail price of data to potential advertisers would be subject to the explicit target group of customers. The origins of the regional spread of values may lie in the differences in attitudes towards disclosing personal information in exchange for targeted services, the ability to use this data by the industry interested in personalized services, regulatory frameworks for consumer protection and/or personal data protection, and macroeconomic factors (level of Internet connectivity, purchase capability, rent, etc.).

Acknowledging the fact that the purpose of this paper is not to establish which variables explain regional differences, it is relevant to check if the "revenues per data record" approach points to such disparities, an effect widely neglected in discussions on the economic value of personal information, and a proof of the utility of this metric for different purposes, as discussed in the conclusion.

Table 2 summarizes the revenues per profile according to world region. The information on revenues is obtained from financial statements, while the number of users per world region is obtained from diverse market analysts, as explained in the corresponding footnotes and, therefore, subject to potential biases. However, the differences between values are sufficiently sizeable to conclude the effect of regional disparities in spite of the existence of considerable potential errors in the data.

Moreover, as revealed, for instance, by recent company communications of Facebook, the increase in the number of accounts is geographically dispersed. For instance, between March 2011 and March 2012, the numbers in the US were growing the slowest and in South America the fastest.[12] Hence, sophistication in terms of estimating the economic value per record based on revenues could be increased still further if such geographical differences were taken into account, of course given that the companies disclose revenue figures regionally disaggregated in sufficient detail.

[12] See: Facebook growth statistics (http://www.internetworldstats.com/facebook.htm) differentiated by world geographical regions.

6. Drawbacks and conclusion

The valuation of individual data profiles and records based on financial figures such as market capitalization and revenue is comparably easy and straightforward given that most companies (publicly traded or not) report such figures. In principle, reporting requirements and accounting standards for public firms mean that data should be not only available for firms whose business model is based on personal information, but also available in a comparable way, allowing the linking of these with information on the number of users/records and thus providing a robust estimate. Moreover, the company's financial results reflect the market worth: the approximated value of data records should reflect the economic value added/revenue generated through the commercial use (or its retail to others for use) within a given market context. In other words, the valuation of personal data relying on a company's accounting figures and/or its financial results can be considered to be a derivative of its true market valuation.

However, there are several challenges related to the use of company-level financial data for approximating the value of personal data. First, a wealth of other company (internal) components beyond personal data influence the firm's financial results, such as its human and physical capital stocks, volume of other intangibles, expertise/know-how, etc. Consequently, approximating the value of personal data by means of market capitalization, revenues and profits could significantly overestimate its value, particularly in industries where personal data are not the main input factor of production. The same holds for approximations relying on the amount of intangible assets and/or derived from the goodwill of a company. Moreover, the financial results of a company can also depend largely on external factors, such as market trends, random shocks and speculation. This means that measures of personal data cannot discard fluctuations over time, following general market sentiments or speculative activity rather than the intrinsic value of the data. Hence the relevance of gathering a set of valuations both over time and cross-sectorial – even regional – to create a baseline. Of course, more elaborate techniques can be used to distinguish to a certain extent general market trends from intrinsic valuations.

Next, the accuracy of data has to be taken with a pinch of salt. A seemingly insolvable problem is that generally accepted accounting conventions give room for choices about what counts as reliable revenues and earnings. This "flexible approach" is still greater for operational assets. The (ac)counting of active users is probably the most serious concern. As stated before, companies can strategically report their number. Even not doing so, it would be extremely complicated especially from the outside of the firm to determine how many profiles are duplicated, belong to a same user or are directly false.

Still more, the relationship between the volume of personal data, its overall value and the financial results of a company may not be linear, not even for companies that overwhelmingly ground their businesses on the commercial use of data. For example, it is likely that the use of personal data in economic activity could include synergistic effects, i.e. for a given company, the economic value of a single isolated data record is lower than the economic value of the same data record in a large, consistent dataset. Such synergies could lead to significant estimation biases.

Finally, the value of personal data, as happens when measuring the value of any other intangible assets, depends on how well they serve the organizations that own them. Therefore, any results presented would be anyway rather sector, market and even time specific, i.e. in a different market niche or moment the value of – the same – personal data might be different.

In spite of the above drawbacks, the paper has tried to put into context with practical figures companies' valuations of the personal information obtained by the preferred approximation of economic returns to data, namely the net present value of revenues per personal record. The availability of a baseline of these valuations would help, therefore, in understanding the evolution of the area, facilitate the comparison of different business models in the same or related sectors and across regions, and increase the general awareness of consumers, policy-makers and society at large of this increasingly relevant new type of asset.

Appendix I. Companies' financial figures per data record (in US$ unless stated otherwise)

Company	Time of reporting[a]	Data records [million; mean]	Intangible assets per data record	Goodwill per data record	Market cap per data record	Revenue per data record	Net income per data record
Facebook	Q1 + Q2 30/06/2012	955	n.a.	0.85	86.00	2.35 (6 months)	0.05
	Q1 + Q2 30/06/2011	710	n.a.	0.19	n.a.	2.29 (6 months)	0.66
	2011[b]	845	n.a.	n.a.		*4.27*	*1.18*
	2010[b]	608			n.a.	*3.25*	*1.00*
	2009[b]	*150*				*5.18*	*1.53*
LinkedIn[c]	2011 (AR)	145	*0.06*[d]	0.08[e]	43.43	3.60	0.08
	2010	90	*0.04*[d]	n.a.	n.a.	2.69	0.17
	2009	55	–	n.a.	n.a.	2.18	−0.07
	2008	32	–	n.a.	n.a.	2.44	−0.14
	2007	17	–	n.a.	n.a.	1.95	0.02
Xing[f]	2011 (AR)	11.7 (0.78)	€0.80[g] (12.00)	€1.92 (28.83)	€19.1 (285.9)	€5.66 (84.9)	€0.80 (12.05)
	2010 (AR)	10.5 (0.75)	€0.68[g] (9.52)	€1.61 (22.56)	€18.3 (256.0)	€5.17 (72.4)	€0.69 (9.60)
	2009 (AR)	8.7 (0.69)	€0.85[g] (10.70)	€0.82 (10.34)	€22.2 (255.1)	€5.18 (65.4)	€−0.20 (−2.46)
	2008 (AR)	7.0 (0.55)	€0.45[g] (5.76)	€1.97 (23.13)	€16.1 (254.5)	€5.04 (64.2)	€1.04 (13.27)
	2007 (AR)	4.8 (0.36)	€0.43[g] (5.71)	€1.93 (25.78)	€47.9 (638.9)	€4.09 (54.4)	€1.17 (15.56)
Experian	31/03/2012 (AR)[h]	*660*[i]	0.69	6.31	£14.22	6. 80	1.09
	31/03/2011 (AR)	*600*[j]	0.71	6.27	£13.07	6.48	1.04

Company	Time of reporting[a]	Data records [million; mean]	Intangible assets per data record	Goodwill per data record	Market cap per data record	Revenue per data record	Net income per data record
Google[k]	2011 (AR)	1727	0.97	4.52	128.65	21.65[l]	5.99
	2010 (AR)	1540	0.74	4.46	136.60	18.34[l]	6.07
	2009 (AR)	1355	0.65	4.12	165.11	16.89[l]	5.47
	2008 (AR)	1173	0.96	4.65	93.00	18.01[l]	4.06
	2007 (AR)	1041	0.50	2.57	242.69	15.76[l]	4.70

Sources: Own calculations from companies' Annual Reports (AR) and financial statements. See the notes below for some assumptions made. Estimated values in *italics*.

[a] End of calendar year (if not stated otherwise). In general, the information provided corresponds to the most recent annual report/quarterly financial statement and is – if possible/needed – complemented by information from previous reports or financial statements (thus gradually going backwards in time).

[b] Various sources; estimated financial figures according to www.cortalconsors.com (no annual report/comprehensive company reporting available yet; first financial report for Q1/Q2 2012 released on 26/07/2012).

[c] Reported accounting figures correspond to GAAP (Generally Accepted Accounting Principles).

[d] Among the intangible assets explicitly considered are patent(s), developed technology, non-competing agreements, workforce in place, and (only in the year 2011) capitalized IPR&D; i.e. no data! Note: The reported amount of tangible assets exceeds the value of intangibles by a multiple. However, capitalizing data stock would be likely to change the corresponding figures fundamentally.

[e] LinkedIn's "goodwill" (similar to the accounting praxis in other companies) is evaluated for impairment annually in the 3rd quarter of the company's fiscal year, and whenever events or changes in circumstances indicate the carrying value of goodwill may not be recoverable. Triggering events that may indicate impairment include, but are not limited to, a significant adverse change in customer demand or business climate that could affect the value of goodwill or a significant decrease in expected cash flows.

[f] In parentheses: number of subscribers to the network (liable to pay).

[g] According to XING AG's annual report, the balance sheet position "intangible assets" (part of non-current assets) comprises the following: software and licenses, internally generated software, goodwill, and "other intangible assets" (see: AR 2011, p. 68). However, on p. 91 it is stated that the "...intangible assets include brands, the customer base, purchased software and internally generated software and goodwill." Brands refer in this regard to a capitalization of the brand Xing (in AR 2011 for the first time). To what extent "customer base" refers to the value of personal profiles and the corresponding data and how and where this is reflected in the balance sheet figures remains somewhat unclear (presumably this is reported as "other intangible assets"). In the table, the reported values for "other intangible assets" per data record are provided. However, in case of the AR for 2011 it was noted that there is an inconsistency in terms of reported values of "other intangible assets" and goodwill; the corresponding values reported on pages 69 and 91 are different.

[h] The accounting year ends on 31 March; i.e. AR 2012 corresponds to the reporting period 1 April 2011–31 March 2012, etc.

[i] The number of records accessible/held by the company is not explicitly reported in the AR (although the "costs of data and data processing" is capitalized). However, the company announced having conducted about 600 million individual and 60 million business credit application and payment history reports in 2011 (i.e. at least this number of profiles were available to the company). Note that "credit reports" is only one of the key businesses of Experian. Hence, there is reason to believe that the total number of data records might be even higher.

[j] Estimated based on the assumptions made for the accounting year 2012 (please note that the company does not provide an explicit number of data records held).

[k] As Google does not report the number of records, a specific estimation had to be prepared: this primarily involved searching for regional statistics of "percentage of Internet users", and then using the Google market share in that region. In general, it has to be noted that Google is quite diverse in its business activities and therefore relating the number of Google users (or search profiles) to the aggregated company figures may imply a certain (higher) bias compared to companies such as Facebook, etc. Direct comparisons of the figures presented across companies should therefore be made with caution.

[l] Only revenues from advertisements (as reported by Google in the ARs) were considered.

Table 3
Annual revenues per data record for an additional set of companies (in US$).

Firm		2009	2010	2011	2012
Yelp[a]	Annual revenues (US$M)	25.8	47.7	83.3	138
	MAU (millions)	23.0	36.0	57.8	81.0
	Revenues per data record (US$)	1.1	1.3	1.4	1.7
Renren[b]	Annual revenues (US$M)	46.7	76.5	118	176
	MAU (millions)	22.0	24.0	38.0	45.0
	Revenues per data record (US$)	2.1	3.2	3.1	3.9
Pandora[c]	Annual revenues (US$M)	55.2	138	274	427
	MAU (millions)	16.0	29.0	47.0	54.9
	Revenues per data record (US$)	3.5	4.7	5.8	7.8
Groupon[d]	Annual revenues (US$M)	14.5	313	1610	2330
	MAU (millions)	0.65	21.4	166	290
	Revenues per data record (US$)	22.4	14.6	9.7	8.0
Zynga[e]	Annual revenues (US$M)	121	597	1140	1280
	MAU (millions)	101.2	216.7	232.8	306.0
	Revenues per data record (US$)	1.2	2.8	4.9	4.2

[a] MAU: http://www.trefis.com/company?hm=YELP.trefis&driver=1086#. Annual revenues: http://www.marketwatch.com/investing/stock/yelp/financials.

[b] MAU: http://techrice.com/2011/04/21/is-renren-seeing-explosive-active-user-growth/. Annual revenues: http://www.marketwatch.com/investing/stock/RENN/financials.

[c] MAU 2009: http://www.statista.com/statistics/190989/active-users-of-music-streaming-service-pandora-since-2009/. MAU 2010 and 2011: http://seekingalpha.com/article/416041-pandora-media-s-ceo-discusses-q4-2011-results-earnings-call-transcript. MAU 2012: http://www.faqs.org/sec-filings/120829/Pandora-Media-Inc_8-K/a12-19115_1ex99d1.htm. Annual revenues: Pandora 2009–2012: http://www.marketwatch.com/investing/stock/P/financials.

[d] Groupon does not report MAU, but total number of subscribers: http://www.trefis.com/company?hm=GRPN.trefis&from=search#/GRPN/n-0240?from=sankey. Annual revenues: http://www.marketwatch.com/investing/stock/GRPN/financials.

[e] MAU 2009–2011: http://www.trefis.com/company?hm=ZNGA.trefis&from=search#/ZNGA/n-2340?from=sankey. MAU 2012: http://techcrunch.com/2012/07/25/zynga-earnings-q2/. Annual revenues: http://www.marketwatch.com/investing/stock/ZNGA/financials.

Appendix II. Annual revenues per data record of a set of additional companies

There has been recently a surge of companies whose business model is based to a large extent in personal profiles. However, few of them display publicly information that can be considered reliable enough. In the following five additional companies to those presented in the main body of the paper are considered, together with the valuation of their revenues per data record and the sources used for the data. Jointly examined with the initial set they present a remarkable coherent valuation range, although it is worthy to insist in this paper not been focused on the details of their business models but in the applicability of the methodology presented.

The first two companies, Yelp and Renren, can be ascribed to the social network type and therefore, they may consistently rank with the initial set. Yelp is a web community of people sharing experiences about local businesses and posting reviews of them. To put in very simple terms, Renren has been nicknamed as the "Chinese Facebook". Pandora, the next case, is strictly speaking "an automated music recommendation service",[13] but somewhat loosely can be defined as a social network based on music and to that regard, collects personal profiles on music preferences (and other) of users. The next two cases, Groupon and Zynga, although relying on personal profiles are considerably further away from the previous examples in terms of business models. Groupon "features discounted gift certificates usable at local or national companies",[14] marketing the appropriate information to interested – profiled consumers. Zynga is a developer and provider of games for both social networking sites and mobile platforms. It uses social network strategies to sell and spread the usage of its games as well as targeted advertising as a main ingredient of its business model.

Table 3 introduces monthly active users of each of these companies, their revenues and, finally, the valuation of revenues per data record from 2009 to 2012. Drawbacks mentioned in the final section of the paper on both estimation of revenues and users are fully applicable here.

References

Berger, P. D., Eechambadi, N., George, M., Lehmann, D. R., Rizley, R., & Venkatesan, R. (2006). From customer lifetime value to shareholder value: Theory, empirical evidence, and issues for future research. *Journal of Service Research, 9*(2), 156–167.

Bond, S. R., & Cummins, J. G. (2000). The stock market and investment in the New Economy: Some tangible facts and intangible fictions. *Brookings Papers on Economic Activity, 2000*(1), 61–124.

Cauwels, P., & Sornette, D. (2012). Quis pendit ipsa pretia: Facebook valuation and diagnostic of a bubble based on nonlinear demographic dynamics. *Journal of Portfolio Management, 38*(2), 56–66.

Choi, W. W., Kwon, S. S., & Lobo, G. J. (2000). Market valuation of intangible assets. *Journal of Business Research, 49*(1), 35–45.

Cohen, J. A. (2005). *Intangible assets: Valuation and economic benefit*. Hoboken: Wiley Finance.

Douplitzky, K. (2009). Le commerce du moi, modèle économique du profilage. *Hermès, 53*, 113–117.

Ehie, I. C., & Olibe, K. (2010). The effect of R&D investment on firm value: An examination of US manufacturing and service industries. *International Journal of Production Economics, 128*(1), 127–135.

Gómez-Barroso, J. L., & Feijóo, C. (2013). Información personal: la nueva moneda de la economía digital. *El Profesional de la Información, 22*(4), 290–297.

Griliches, Z. (1981). Market value, R&D and patents. *Economic Letters, 7*(2), 183–187.

Grosu, V., Hlaciuc, E., Mates, D., Bostan, I., Socoliuc, M., & Mihalciuc, C. (2011). Accounting assymmetries regarding the measurement of intangible assets. In K. S. Soliman (Ed.), *Proceedings of the 17th IBIMA Conference on Creating Global Competitive Economies: A 360-Degree Approach* (pp. 761–773). Milan: IBIMA.

Gu, F., & Wang, W. M. (2005). Intangible assets, information complexity, and analysts' earnings forecasts. *Journal of Business Finance and Accounting, 32*(9–10), 1673–1702.

Gupta, S., & Zeithaml, V. (2006). Customer metrics and their impact on financial performance. *Marketing Science, 25*(6), 718–739.

Hall, B., & Oriani, R. (2006). Does the market value R&D investment by European firms? Evidence from panel data of manufacturing firms in France, Germany and Italy. *International Journal of Industrial Organization, 24*(5), 971–993.

Hulten, C. R., & Hao, X. (2008). *What is a company really worth? Intangible capital and the 'market to book value' puzzle. NBER Working Paper No. 14548*. National Bureau of Economic Research. Available at. http://www.nber.org/papers/w14548

Hunter, L., Webster, E., & Wyatt, A. (2012). Accounting for expenditure on intangibles. *Abacus – A Journal of Accounting Finance and Business Studies, 48*(1), 104–145.

Kumar, V., & Shah, D. (2009). Expanding the role of marketing: From customer equity to market capitalization. *Journal of Marketing, 73*(6), 119–136.

Kuneva, M. (2009). *Keynote speech at the Roundtable on Online Data Collection, Targeting and Profiling. Reference: SPEECH/09/156. Brussels, 31 March 2009*. Available at. http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm

Lev, B. (2003). Remarks on the measurement, valuation, and reporting of intangible assets. *Economic Policy Review, 9*(3), 17–22.

Palmon, D., & Yezegel, A. (2012). R&D intensity and the value of analysts' recommendations. *Contemporary Accounting Research, 29*(2), 621–654.

Pavlou, P. A. (2011). State of the information privacy literature: Where are we now and where should we go? *MIS Quarterly, 35*(4), 977–988.

Potoglou, D., Patil, S., Gijón-Tascón, C., Palacios, J., & Feijóo, C. (2013). The value of personal information online: results from three stated preference discrete choice experiments in the UK. In *European Conference on Information Systems, ECIS 2013* Utrecht, The Netherlands, (pp. 1–12). Retrieved from. http://www.staff.science.uu.nl/~Vlaan107/ecis/files/ECIS2013-0899-paper.pdf

Schulze, C., Skiera, B., & Wiesel, T. (2012). Linking customer and financial metrics to shareholder value: The leverage effect in customer-based valuation. *Journal of Marketing, 76*(2), 17–32.

Whitwell, G. J., Lukas, B. A., & Hill, P. (2007). Stock analysts' assessments of the shareholder value of intangible assets. *Journal of Business Research, 60*(1), 84–90.

Wyatt, A. (2005). Accounting recognition of intangible assets: Theory and evidence on economic determinants. *The Accounting Review, 80*(3), 967–1003.

Claudio Feijóo holds a M.Sc. and Ph.D. in Telecommunication Engineering and a M.Sc. in Economics. Currently he is full Professor at Technical University of Madrid where he researches on the future socio-economic impact of emerging information society technologies, in particular, from a mobile and/or content perspective. From his work experience, he enjoyed very much his two years at the Institute for Prospective Technological Studies of the European Commission, the direction of the Chair in Telecommunications Regulation and Information Society Public Policies at Technical University of Madrid, and the university spin-off company he helped to launch and manage. He lectures regularly in seminars and postgraduate courses and has authored many publications in books, and main journals and conferences.

José Luis Gómez-Barroso is an Associate Professor in the Department of Applied Economics and Economic History at Universidad Nacional de Educación a Distancia (UNED). He holds a Ph.D. and a degree in Economics from UNED. He also holds a degree in Telecommunication Engineering from the Universidad Politécnica de Madrid as well as another degree in Law from the Universidad Complutense. He is a "PURC-Senior Research Associate" (University of Florida) and has been a European Union Fulbrighter Visiting Scholar at the Columbia Institute for Tele-Information (CITI), Columbia University. He has published more than sixty academic papers and chapters in books.

Peter Voigt graduated in economics at Martin Luther University (MLU) Halle-Wittenberg in 1998; worked then as research associate for IAMO (1998–2006); defended his Ph.D. thesis in 2004 at MLU; had a post doc Marie Curie Fellowship at INGENIO (2003–2005), worked as Scientific Officer and Advisor for Economic & Policy Affairs for European Commission (JRC-IPTS, 2006–2012), was Visiting Scientist at University of Barcelona (2012), associated expert at Polytechnic University of Madrid (CeDInt, 2013) and Senior Consultant at Worldbank – IDB (2013). Currently, Mr. Voigt is Senior Researcher/Project Manager at IAMO.

[13] http://en.wikipedia.org/wiki/Pandora_Radio.
[14] http://en.wikipedia.org/wiki/Groupon.