D3.5

Socio-economic value of personal information

V1.0 / 2015-04-01

Jonas Breuer (iMinds), Rob Heyman (iMinds), Shenja van der Graaf (iMinds)

This report presents a socio-economic perspective on the tool for user-centred personal data management as envisioned by the USEMP project. It evaluates how the ecosystem can evolve in the future shaped by the USEMP tool, identifying business opportunities and challenges that can be seen to arise when the user has the means to assert more control. In doing so, the deliverable focuses on the industry perspective. The industry of online advertising was chosen as the use case of investigation. Combining an in-depth assessment of the current situation in the form of a state-of-the-art with empirical data gathered through expert interviews, this report draws out an existing value network as well as provides some recommendations on how the USEMP tool can make most sense in socio-economic terms.

Project acronym	USEMP
Full title	User Empowerment for Enhanced Online Presence Management
Grant agreement number	611596
Funding scheme	Specific Targeted Research Project (STREP)
Work program topic	Objective ICT-2013.1.7 Future Internet Research Experimentation
Project start date	2013-10-01
Project Duration	36 months

Workpackage	WP3
Deliverable lead org.	iMinds
Deliverable type	Report
Authors	Jonas Breuer Rob Heyman Shenja van der Graaf
Reviewers	Ali Padyab (LTU) Theodoros Michalareas (VELTI)
Version	1.0
Status	Final
Dissemination level	PU
Due date	2015-03-31
Delivery date	2015-04-01

Version	Changes
0.1	Table of contents
0.2	State of the art, methodology
0.3	Finalization state of the art, empirical research
0.4	Results
0.5	Scenarios, conclusion
0.6	Full draft for review
0.7	Integration of the first reviewer’s comments
0.8	Integration of the second reviewer’s comments
1.0	Final

Table of Contents

List of Figures and Tables 2
1. Introduction 3
1.1. Purpose and Scope of Document 3
1.2. Structure of the Deliverable 4
2. State of the Art 5
2.1. Business Ecosystems and Value Networks 5
2.2. Current Personal Data Value Network 6
2.2.1. Actors, their Roles and Relationships 7
2.3. State of Legislation: The GDPR 15
2.4. A Note on Privacy in the USA 18
2.5. Existing Tools and Initiatives 18
2.6. Value Propositions 21
3. Methodology 23
3.1. Research Question 23
3.2. Business Models and Value Networks 24
3.3. Expert Interviews 25
3.3.1. Interviews 26
3.3.2. Expert selection 27
3.4. Scenario Development and Workshop Iterations 28
4. Results of Empirical Research 29
4.1. Ecosystem Description Based on Interviews 29
4.1.1. Value Network: Actors, Roles, Relations 29
4.1.2. User-Centric Data Management 36
4.1.3. GDPR 37
4.1.4. Data 38
5. Scenarios for the Instrumentalisation of the Shared Information Value 42
5.1. As-is and As-would-be 43
5.2. USEMP: A Centralised Tool 45
6. Conclusion 47
Bibliography 49
ANNEX A) Interview Questions 51
ANNEX B) Contacted experts and elite 52

List of Figures and Tables

Figure 1: The Value Network of Online Advertising 6
Figure 2: Real-Time Bidding (Source: (Siebelink, 2014) 9
Figure 3: Datalogix enrichment of offline buying pattern for Facebook Targeting Criteria 11
Figure 4: LinkedIn's, Facebooks & Twitter's Revenues and MAUs (Source: Heyman, PhD in progress) 12
Figure 5: Facebook Audience Network 14
Figure 6: Datacoup's Data Profile 20
Figure 7: Where to locate the tool in the Value Network 43

Table 1: Deeplinking 14
Table 2: Business Model Matrix (Source: Ballon, 2007, p.10) 25

1.Introduction

1.1. Purpose and Scope of Document

USEMP is a multidisciplinary research project, integrating the perspectives of lawyers, engineers, computer scientists, marketing experts and social scientists. The project aims at developing a framework that will empower users by enhancing their control over the data they distribute or interact with. At the core of this lies the idea to reduce the existing asymmetries of processing and control means between companies and citizens. In this context, the goal of Task 3.7 is to contribute a socio-economic perspective on shared information value, i.e. to assess the actual value of personal data. No commonly accepted methodology for estimating this value exists. According to the Description of Work of the project and the task, approaches tend to rely either on individual user perceptions of the value of their personal data, or market valuations thereof. The reason for this is that value is (always) subjective, i.e. it only arises in relation to a specific person, company, an objective, need etc.

The task at hand evaluates how an ecosystem as envisaged in USEMP can evolve in the future, identifying business opportunities and challenges that can be seen to arise when the user has the means to assert more control. This is particularly relevant for USEMP, as it determines its sustainability. The relevance of Task 3.7 thus lies beyond the project, as a user-centric data management initiative cannot meet its objectives if it is not sustainable in the mid- to long-term.

This deliverable elaborates on the different values of personal data. It focuses on the industry perspective. It was chosen to do so, because the envisioned tool cannot be successful without backing from the industry. Whereas user requirements are specifically taken into account by the project, requirements of affected businesses are still missing or misunderstood. The industry of online advertising was chosen as the use case of investigation. This decision was taken against the backdrop of USEMP's initial scenario concerned with "economic value awareness" (see USEMP D2.1) This also makes sense in general, as this $120 Billion economy (The Economist, 2014) is inseparable of personal data, i.e. information about users and customers; for profiling, targeting and communicating with relevant advertising. It is the most advanced and most influential market place for personal data. As the industry depends on the use of such data, it is also likely to be inclined to facilitate developments so as to be able to continue using it.

This deliverable does also not aim at reducing or ceasing data re-use per se. It rather evaluates an ecosystem in which data can be used efficiently and in accordance with the requirements of users. This is the case because "the increased quantity, value and importance of personal information and the ubiquity of mobile platforms has rendered inadequate traditional privacy practices." (van der Graaf and Clippinger, 2012) The main objective should be to create mutual benefits for all stakeholders. According to the World Economic Forum one needs to distinguish between "user-centred personal data ecosystems, where all stakeholders are focused on empowering individuals" (Word Economic Forum, 2014, p. 3) and "user-centric personal data ecosystems, where individuals themselves are solely responsible for managing the use of all data related to them." A trust framework "that specifies all duties of relevant stakeholders" (van der Graaf and Clippinger, 2012) is likely to

enable to combine the idea of a user-centred personal data ecosystem with the interests of the industry.

1.2. Structure of the Deliverable

Chapter 2 'State of the Art' (Sota) provides the necessary background to understand the ecosystem of display advertising and the value-creation in relation to personal data. The Sota is an overview of the status-quo. It commences with an introduction to the state of the art of business ecosystem and value network research, succeeded by an overview of the current value network of personal data re-use in the field of display advertising. This task is central, as the ecosystem is mapped and described, and a simplified value network provided. This is complemented by an introduction respectively of the state of legislation, of privacy in the USA[1], of existing value proposition and tools. This is necessary in order to identify the ways in which the proposed USEMP business ecosystem could evolve best in the future. The material for this first part of the deliverable was gathered through desk research.

Chapter 3 introduces the guiding methodology of this deliverable, describing the research question, value network research, expert interviews, and the constructing of scenarios.

Chapter 4 describes and evaluates the empirical research. Several actors and stakeholders within the defined value network were interviewed. They provided indispensable insights on their operations and roles as well as on the status quo of the personal data market in general.

Chapter 5 builds upon the value network, and all additional information gathered. It constructs and evaluates different potential futures of the personal data market with a focus on Europe. These scenarios describing potential futures are developed and improved in an iterative fashion. In this way, a value network could be constructed and its most important aspects could be qualitatively evaluated. The iteration approach promises to provide the most realistic outcomes in line with the needs of real-life stakeholders, prioritising economically and strategically viable scenarios based on mutual benefits.

[1] A short introduction to the situation in the USA was deemed necessary. This is the case because a major share of innovation and businesses that shape the personal data market originate there.

2.State of the Art

2.1. Business Ecosystems and Value Networks

In the field of Ecology the term ecosystem describes a biological community of interacting organisms and their physical environment (Moore, 1993). In general use, it indicates a complex network or interconnected system, entailing that linear dependencies (as in traditional value chains) make way to multi-directional dependencies. We adapt the definition of a business ecosystem as a network of actors that are interrelated and mutually dependent, the ecosystem being defined and delimited by its relation to a common interest (MOORE, 1993). Its strength and weakness depends on how each actor contributes ('adds value') to the ecosystem (Ballon, 2007). Interactions are key, and relationships describe the interactions between actors within the ecosystem. They co-evolve over time with actors and their roles, are influenced by the wider environment as well as by each actor's objectives and interests.

Most business model frameworks proposed in the literature, notably Osterwalder (2010) and Chesbrough (2002), are, however, more suited for individual firms than for guiding collective innovation processes. It is therefore necessary to consider a stream of research that provides a coherent treatment of the most relevant business model parameters while at the same time emphasising relationships between actors in an ecosystem setting. Ballon (Ballon, 2007), Bouwman (Bouwman, Haaker, & Vos, 2005) and Allee (Allee, 2009) thus propose to enrich the classic value chain approach in changing environments such as telecommunication and online services because, here, value is no longer created in a linear fashion. Their research includes multiple companies creating value through a web of interconnected value chains associated with a similar value proposition. The origin of this approach to business modelling is arguably the internet-based economy (Al-Debei & Avison, 2010; Hawkins, 2001), where novel ways of interacting with customers and within networks, have become a source for innovation and success.

This is also the case for online advertising and the market online personal data re-use in general. As mentioned in the introduction, online advertising is a $120 Billion economy that is inseparable of information about users and customers; for profiling, targeting and communicating with relevant ads[2]. It is the most advanced and most influential market place for personal data. It is also highly complicated, as the ecosystem is a network of interrelated and mutually dependent actors. All of these actors contribute value to the ecosystem, which has evolved far from a linear value chain. The chain between advertiser and public or users has become more complex due to the advent of several intermediaries such as advertising agencies, demand- and supply-side platforms, ad exchanges, ad networks, ad servers, buying solutions, analytics and many more. In this context it is most important to realise that the ecosystem is reshaping itself according to and in line with available technology as much as business development potential.

[2] "Ad" is the abbreviation of "advertisement". Both will be used interchangeably from here onwards.

2.2. Current Personal Data Value Network

The ecosystem of display advertising is founded on an interdependent, two-sided structure, where advertisers constitute the one side, publishers the other, and several entities in between that facilitate the interactions. The whole ecosystem increasingly depends on data (not only personal data) and the actors in between exist to make interactions more efficient, also based on data. The division persists mainly because publishers want to sell ad inventory at the highest possible price while advertiser need to buy it at the lowest possible price.

Figure 1 depicts and simplifies the value network; green boxes represent actors, the light grey boxes within their roles; arrows stand for relationships and interactions between actors or groups of actors. The grey background boxes with dotted lines represent these actor groups.[3] The following section describes the value network in more detail. It is structured according to the historic evolution of the ecosystem and its structural actors. It can be argued that each new actor emerged to solve a problem caused by the previous actor. Today's online display advertising network is thus an incremental value network where each actor has built on shortcomings of the previous actor. While each new actor tried to reduce complexity, overall complexity increased, with an increasing number of intermediaries between publisher and advertiser.

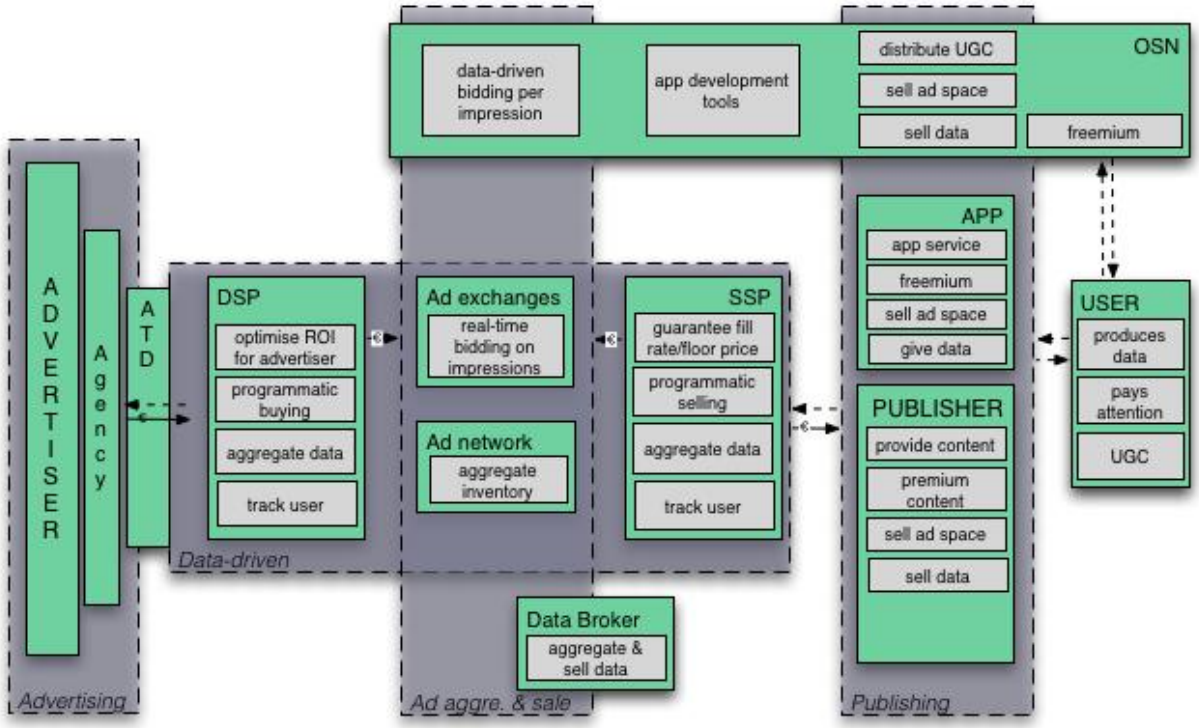

Figure 1: The Value Network of Online Advertising

[3] To increase readability of the figure, not all arrows are drawn out. This means that the actors have more complex relationships in reality.

2.2.1. Actors, their Roles and Relationships

Users

The user is the actor in the value network around which all other actors' activities ultimately revolve. However, the user is passive, providing personal data and User Generated Content (UGC), eyeballs/attention in return to impression, and money when buying products or consuming services. He/she interacts in the first place with content provided by the publisher (mobile application or online content) and related ads or with an Online Social Networks (OSN) (same attention as for publisher, but also UGC). Then, with advertiser or company that is offering a product/service for sale: visiting advertiser in case ad is interesting, proceeding through the sales funnel and ideally reaching its end, the purchase. The advertiser is also collecting information. Notably, data intensive operations seem to take place "far away" from users.

Publisher

More traditional publishers such as online newspapers are grouped with mobile apps.

Publishers are in the business of providing inventory for advertisers in the form of online content or mobile applications. This group online content providers (e.g. newspapers) and includes mobile apps, because they are also selling ad space and building their business models on advertising; either in a "free" business model where the online content or application is offered for free to the end-consumer in exchange for advertisements revenue or in a freemium model where there is a "free" version of the service related to advertisement and additional premium features that the consumer has to pay for to access). Publishers are in direct contact with the user, providing content and receiving 'eyeballs' in return. Mobile applications are especially important today, as they add the dimension of location services. Here, location data can be gathered and re-used for targeted and context-specific ads, thus presenting another source for personal data leakage.

In the direction of the advertiser, the publisher sells ad space, but also provides certain kinds of data mainly for targeting (especially context of impression). The publisher pays supply-side platforms (SSPs) for their services, provides information for them to increase efficiency of services through aggregate data. Most importantly, SSPs guarantee to the publisher a floor price, i.e. a minimum price for an ad space. Furthermore, there are interactions with ad-exchanges (providing impressions, receiving CPC[4], CPM[5] or CPA[6]), ad-networks (providing remnant inventory and receiving CPM) and increasingly also OSNs: the publisher receives traffic through OSNs, but is also in competition as OSN is taking advertising space and eyeballs.

Advertiser

The counterpart to the publisher is the advertiser, who provides the advertisements to be displayed on/into the publisher's content. The advertiser may or may not act through an

[4] CPC describes the cost per click

[5] CPM or cost per mille is the cost of a 1000 delivered ad impressions. Other performance based compensation modes are e.g. cost per engagement (CPE) and cost per view (CPV).

[6] CPA or cost per action is the cost for a delivered result or action from an advertisement campaign, a very popular variant is CPI, cost per installation for advertisement campaigns that result to installations of mobile applications

advertising agency that helps to generate and place the ad. Advertisers share data with demand-side platforms (DSPs) and supply-side platforms, ad-exchanges and ad-networks, on target groups and users who have previously been noted as visitors/buyers. They pay the publisher of the ad and the intermediaries for facilitating the process. Due to the complexity of the value network and its deviating value chains, advertisers need sufficient information: about which actor is performing which role, and how successfully they do so. This is also a reason for the significance of user IDs, as they are necessary for conversion attribution.

Ad Network

The ad network is the oldest aggregator of publishers and their impressions. It buys remnant inventory publishers had difficulties to sell, and sells in CPM. This actor has come about in the nineties, as the number of online publishers increased. As a result advertisers or advertising agencies had difficulties planning and buying from the increasing number of publishers. Ad networks buy remnant ad inventory[7] from publishers. Then the inventory is categorised and sold through the ad network. Because the ad-network buys inventory in advance from publishers, publishers may not know for what price their inventory is sold and to whom.

Advertisers approach ad networks with specific aims, such as X number of impressions for Y amount of time and a particular budget. The ad network tries to combine its inventory to meet all advertisers' goals. Also the advertiser may not know from whom it is buying or what the ad network paid for it. Transactions are most often in CPM. The ad network is a black box that reports back to advertisers after the campaign. Since pre-bought inventory might not be used, it can be resold to other ad networks, adding to the no-transparency. As a result of the inefficiency and non-transparency, ad networks are becoming increasingly redundant today's data-driven environment.

Ad exchanges and real-time buying

Ad exchanges are an answer to shortcomings of ad networks; a bidding platform (like a stock exchange) where impressions are sold & bought. They are more transparent, because they function in real-time, selling per impression on a real time basis. Information is provided about the context of each impression. That transactions are made in real-time means that impressions are bought in the time period a website loads: a 'bid manager' has to decide whether or not to buy the impression. If yes, a bid is issued and the winner gets its ad served. A bid manager is a piece of software that works on behalf of the advertiser, deciding to bid based on the advertising goal and budget. See the next Figure.

[7] Ad inventory is space on web pages suited for advertisements, the most exemplary format is the banner although different formats exist. An overview of ad formats can be found here http://www.iab.net/guidelines/508676/508767/ad_unit.

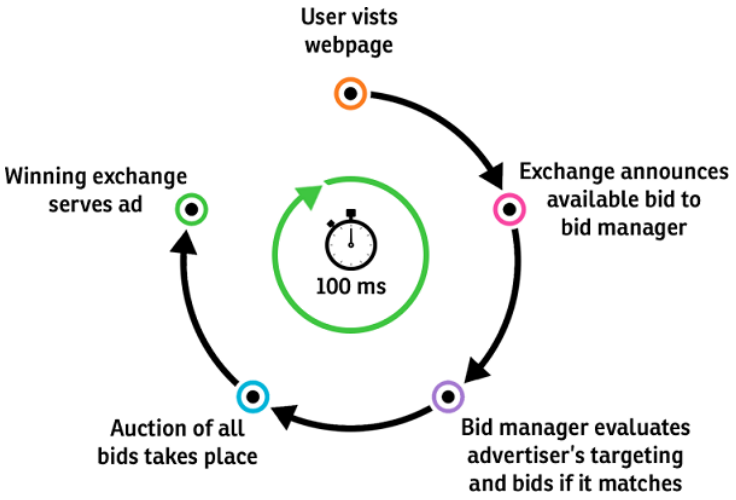

Figure 2: Real-Time Bidding (Source: (Siebelink, 2014)

This process exhibits decisive advantages for both publishers and advertisers. Firstly, there is no wasted inventory for publishers or ad networks[8]. Secondly, publishers receive a larger profit because the cost per impression is higher as advertisers are outbidding each other. Also, publishers can often define a minimal value or floor value or decide to block certain advertisers. Lastly, advertisers or advertising agencies are more in control because they receive real-time feedback on their impressions, which also enables them to pick certain publishers over others.

SSP and DSP: Programmatic buying

Automation is among the most decisive developments for the ecosystem, best signified in programmatic buying and its most important function of real-time bidding (RTB). Ad exchanges have increased transparency through real-time bidding, but the amount of information to evaluate a Return On Investment (ROI) has gone beyond human capacity. Multiple impressions are bought of multiple publishers. This increased the need for an integrated platform to bid on behalf of the advertiser. As a result, two kinds of platforms exist, representing the SSP and the DSP. The former intends to increase revenue for publishers, while the latter bid for a price as cost-efficient as possible for advertisers.

The SSP enables publishers to manage inventory and maximize revenue, through an efficient, automated and secure technology. Through SSPs, publishers are able to sell ad space to advertisers, but also to know about the various income streams and audiences. Floor price and fill rate are decisive in this regard. The former is a limit set for publishers on how low a price can be charged for inventory in the programmatic buying process. The latter determines the rate at which a publisher has successfully sent and received a request for an ad impression. The SSP usually intermediates between the publisher and the ad exchange, which then interacts with the DSP on the advertiser side. Still, SSPs might also receive offers directly from advertisers, DSPs, and ad networks.

DSPs decide in real-time what inventory is best bought to serve the needs of their advertiser and the attributes of the desired audience. Through data, DSPs predict the behaviour of particular audiences that are sold for particular prices. It tries to predict where the most cost-

[8] Ad networks can sell to ad exchanges as well.

efficient audience is, at a given time. Different forms of data are important for online display advertising Based on aggregated data, they determine the best ROI, corresponding to the objectives of the campaign of an advertiser. For the programmatic buying, on the other hand, DSPs need to track data for individual user-IDs so that they can attribute the conversion of a user to viewing a promotion or taking an action to a specific ad impression (conversion attribution).

Agency Trading Desk (ATD)

Agency trading desks[9] are DSPs owned by an advertising agency. They are an answer to the loss of control due to the appearance of ad exchanges and then DSPs. “Agency trading desks are centralised, service-based organisations that serve as a managed service layer on top of a DSP. They originated from trading desks inside advertising agencies that were tapping into auction-based advertising exchanges.” (ANA, 2011, p. 5) Advertisers are in danger to pay double, to agency and ATD as their belonging together might not be clear. The so-called markup fee that ATDs charge is not made transparent because it would make them less competitive.

Data Broker

Data brokers aggregate data into meaningful profiles for advertising purposes. Data can be gathered through owned panels and surveys, or obtained and combined through partnerships. The most notorious example is the partnership between Facebook and the data brokers Axciom, Datalogix and Epsilon, which allowed Facebook to attach new information to its users profiles (Facebook, 2013). The process consists of two steps. Resulting information or tags depend on what the first party buys from the broker. First, the data broker and the database of a first party look for a common identifier (e.g. email address). A common record can also be a HTTP cookie. Then, the data broker attaches its information to the common records. In the example below, Facebook enriched its database by adding food purchase patterns purchased at Datalogix.

[9] An overview of ATBs: Havas - Adnetik (became independent in 2010); IPG - Mediabrands Audience Platform; MDC - Varick Media Management; Omnicom - Accuen; Publicis - Audience on demand; WPP - Xaxis

Figure 3: Datalogix enrichment of offline buying pattern for Facebook Targeting Criteria

OSN

OSNs are different from other actors. First, they might perform roles that very similar to other actors in the value network; to DSP and ad exchange (profiling, targeting) or to publisher (showing ads). Second, they evolve from a single publisher into an ad exchange where additional publishers are added to the impressions inventory they server ads to; Facebook, Twitter, LinkedIn etc. have first focused on user growth and later on monetisation through data. Social media have an advantage in data in terms of kinds of data and means to coerce consent form data subjects to gather and monetise their data.

User growth. The move from prioritising user growth to monetisation has come about as UGC is part of the definition of social media itself. Social media are built on the technological and ideological foundations of web 2.0 (Kaplan & Haenlein, 2010), using UGC to attract users to services in order to support and expand a loyal user base. Social media affordances are engineered with input from user generated content and data to lure users in (Cohen, 2008) or to create a basis for ongoing connectivity (Van Dijck, 2013). This is also illustrated by the relative long phase between the founding of an online social network (OSN) and their IPO[10]: LinkedIn launched in 2003 and filed their IPO in 2011, Facebook launched in 2004 and filed the IPO in 2012, Twitter launched in 2006 and filed the IPO in 2013. In this regard we can see that each OSN that filed an IPO had a period of user growth paid with venture capital. Monetisation of data usually intensifies one year before the IPO[11] to convince stockholders, that the platform is financially self-reliant and profitable.

Data monetisation. There are at least three means to monetise data, directly or indirectly; advertising, premium and granting access to data. All have been summarised in the figures below. Notice how user increase (monthly active users (MAUs)) has stabilised while revenue streams have increased.

[10] IPO refers to “Initial Public Offering”. This is the stock market launch of a company, in which shares of stock are offered to the public.

[11] At first sight this makes sense, an IPO-filing contains public information with regard to revenue streams.

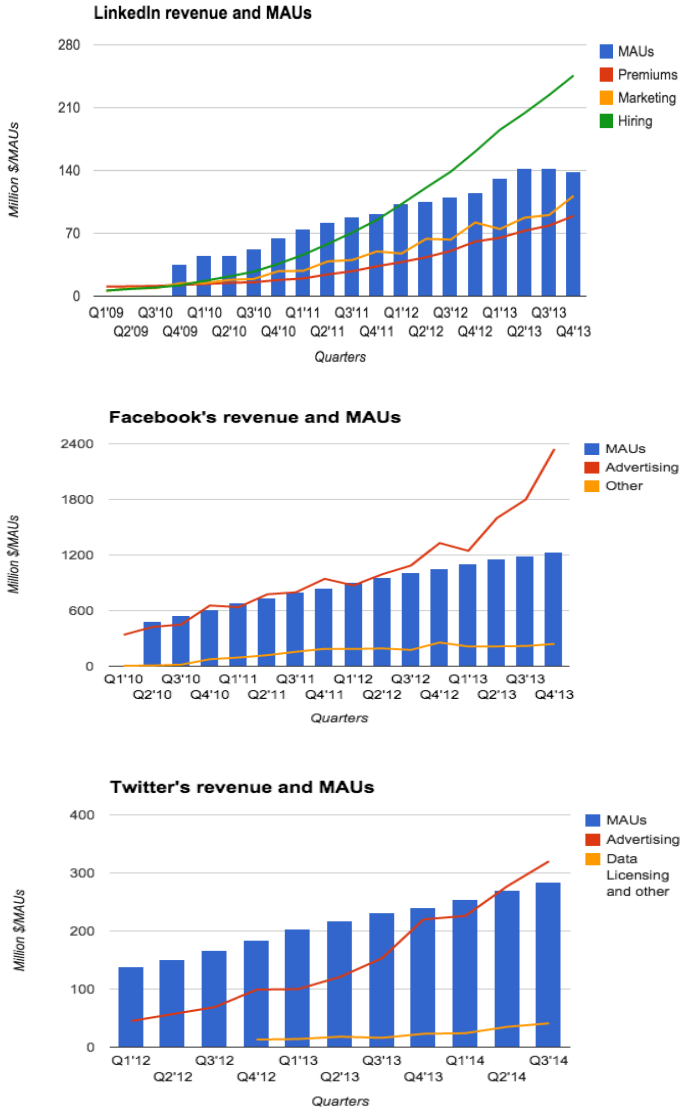

Figure 4: LinkedIn's, Facebooks & Twitter's Revenues and MAUs (Source: Heyman, PhD in progress)

This monetisation trajectory coincides with an increase in ad units and thus possibilities to serve ads. On Facebook and Twitter an increase of advertising could be felt as more and more sponsored content appeared inside the UGC stream as sponsored stories, pages and posts for Facebook, starting in January 2011; sponsored tweets, tweeps and trends on Twitter from April 2010 onwards (Heyman, PhD in progress).

Also premium is a model for monetisation. LinkedIn is an example, offering premium accounts for job seekers and hirers. In this case user data is sold indirectly because premium users receive access to more complete information.

Lastly, data may also be used directly, LinkedIn and Twitter monetise their data directly. Twitter sells access to its Application Programming Interface (API) so that companies can analyse Twitter's 'Firehose' of all tweets in real time. LinkedIn introduced its Audience Network, where profile information was used to target advertisements outside its own platform. It worked together with Collective Media, a company that allows audience targeting on ad networks: "Whenever someone visits LinkedIn, a cookie will be placed on their browser, which will identify them as a LinkedIn member when they visit a partner site. Personally identifying information will be removed, but members will be grouped into

different, targetable categories. As with Yahoo and Google's similar ad-network targeting, anyone will be able to opt out of this program." (Schonfeld, 2008)

Further data expansion. The increase in so-called horizontal data points has become important. Ad networks and ad exchanges have access to a wide selection of data points based on which users can be found. But they can only have information about the behaviour of visitors throughout the sites they track and about the content of each site. Facebook now also uses this data through conversion pixels, social plugins and like-buttons. Each time one of these items loads, Facebook can trace users through cookies. The expansion or privacy invasion is leveraged by its integration in everyday life. Users of OSN are presented with this expansion via a take it or leave it approach. OSN announce an expansion of data use and users have to decide by leaving or staying on a service.

Mobile. Lastly, Twitter and Facebook are expanding towards the new mobile market by acquiring companies to build mobile advertising exchanges. These exchanges leverage the already present data from within the OSN (for example, profile information and UGC), horizontal data gathered from social media plugins and lastly mobile data through owned and third party applications. This allows for an integrated marketing approach on multiple devices.

Recent technological developments and acquisitions of mobile advertising companies have considerably enhanced Facebook and Twitter's mobile tracking capabilities. Facebook is now able to track who installed which application, but more importantly what users do with other applications installed on their mobile devices; this enables the company to compose a detailed dossier of the user (Fowler, 2014). From a mobile marketing perspective this is an important breakthrough because it allows targeting within apps and mobiles. This was previously impossible because cookies did not work on mobile devices. It is relevant to discuss this new development to understand how data is shared between multiple 'family' companies[12] within Facebook and how tracking practices have expanded beyond Facebook.

Mobile in-app advertisements are primarily used to drive app installs. This makes sense because they advertise a good that can be bought with a mobile device. This increases the relevance of advertisements on the one hand, but it also increases the ease to track conversions. This is so because the purchase is made with the same device. This revenue is considerable for Facebook and Twitter: "The ads contributed 'real revenue' to its bottom line said Mark Zuckerberg, and helped total ad revenue from mobile jumped from 23% to 30%. Meanwhile, app install ads had helped Twitter file to IPO with over 65% of ad revenue coming from mobile. By late 2013, the world had proof that app install ads were poised to become a giant business." (Constine, 2014)

Tracking users' behaviour in advertisements is a requirement for targeting and attribution[13] purposes. As mentioned above, mobile tracking is difficult because cookies do not work in apps or on mobile browsers. This means that a new standard or protocol has to be devised in order to implement tracking purposes. Facebook and Twitter have acquired companies to create their own development environment and advertising infrastructure. This way they

[12] Facebook Payments Inc., Atlas, Instagram LLC, Mobile Technologies Inc., Onavo, Parse, Moves, Oculus, LiveRail and WhatsApp Inc.

[13] Attribution is used to refer to the party or parties that convinced a potential client to buy the product. In digital marketing it requires a degree of tracking to attribute the sale to particular parties.

control these protocols entirely. A short summary of this development is provided to illustrate the need for personal data flows between Facebook's companies.

In 2013, Facebook has bought Parse[14], a company that provided an environment to let app developers build their applications more easily. Parse offers managing and testing possibilities next to the addition to integrate with Facebook's API and News Feed to increase organic sharing. But this is not the only reason to coax application developers. Facebook launched its Facebook Audience Network in April 2014. This allowed application owners to sell their advertising space to Facebook's audience network (Facebook, 2014).

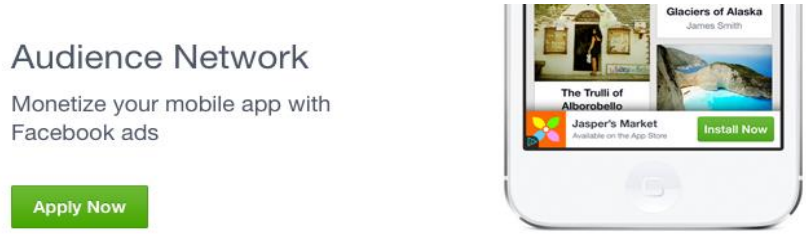

Figure 5: Facebook Audience Network

Deeplinking. This arms race brings us to new tracking technologies for mobile applications applied by Facebook, Twitter and independent companies such as Tune[15]. So-called deeplinking allows application developers to add lines of code within applications that act as touchpoints. These touchpoints work in two directions; they can be used to refer a user to a particular location or they inform a party that a user came from a particular location. For example, before deeplinking, it was impossible to refer to an application or a particular part of an application. So if someone wanted to show a hotel on TripAdvisor, it had to pass via the browser because an URL is able to refer to a particular location: tripadvisor.com/particularhotel1. Now application developers can tag parts of their app with similar URL's or deeplinks as shown in table 1, below.

Developer	**Deeplink – e.g.**	**Purpose**
Twitter	twitter://timeline	Opens the Twitter app and links to the user's timeline
Facebook	fb://profile	Opens the Facebook app and links to the user's profile
Yelp	yelp://	Opens the Yelp app (note: this example does not include any routing parameters)

Table 1: Deeplinking

In conclusion, the integration of Parse and deeplinking implies that HTTP cookies are no longer the only means to track users. More importantly this allows Facebook to track users through their mobile apps via every app that relies on Parse and use of Facebook or Parse

[14] https://www.parse.com/ .
[15] http://www.tune.com/.

SDK (Software Development Kit). This data is part of Facebook Audience Network, which is leveraged by another Facebook company, Atlas[16].

Groups

The value network in Figure 1 is constituted by several broader groups of actors and roles. These groups overlap and some actors belong to more than one. The figure mainly depicts the very basic interactions between groups. For example, members of the *Data-Driven* group pay members of the *Ad Aggregation & Sale* group for the real-time bidding on impressions, which latter group facilitates. However, individual actors also interact among each other: the SSP receives bids for ad space from ad exchanges and networks, but it can also receive them directly from advertisers and their partners.

The *Advertising Group* in the value network is the most clearly delimited one. Advertisers carry out respective roles. They are possibly but not necessarily supported by agencies and their ATDs.

The two groups of *Ad Aggregation & Sale* and *Data-Driven* overlap most significantly. Together, they constitute the core of the modern system of display advertising: programmatic buying and its backbone of real-time bidding. With programmatic buying, the buyer sets parameters (budget, bid price, goals, attribution model, network reach…) based on campaign settings, which are applied to behavioural or audience data.

Ad Aggregation & Sale includes a potential role of an OSN (data-driven bidding for impressions) in addition to ad exchanges, ad networks and also certain roles of data brokers. We argue that this group's activities revolve primarily around targeting of users, i.e. they are primarily concerned with data describing data subjects.

The *Data-Driven* group is more concerned with what we call evaluation, i.e. data which describes or predicts ROI. In terms of ROI, customer value lifetime (predicting total revenue from one customer) and conversion attribution are decisive. It consists of the DSPs and SSPs, and overlaps with former group by also including ad exchanges and ad networks. All members of this group may buy from publisher via intermediaries, but also directly.

It is noteworthy that OSNs are, despite being relatively new actors in this business ecosystem, in a position to occupy several of the roles. Other actors traditionally occupied these roles. A reason for this is the access to diverse (and comprehensive) data concerning their users. User data being an essential asset in the personal data market and beyond, OSNs obviously occupy a strong position. Thus, the value network figure depicts that the OSN adopts some publishing roles. As traditional publishers, it interacts with users through ads, which are shown on its websites, and it sells some kind of data: where an ad is shown, in what context, and who is the target group.

Next, we provide some insights into the current state of legislation.

2.3. State of Legislation: The GDPR

The proposed General Data Protection Regulation (GDPR) intends to ameliorate current, fragmented data protection directives dating back to 1995 in the EU. It is not our goal to

[16] http://atlassolutions.com/

summarise the whole regulation here, as other deliverables of work package 3 serve that purpose. Deliverable 3.1 discusses Fundamental Rights Protection by Design in this light. Deliverable 3.4 addresses the general coordination of legal aspect in USEMP. The following section introduces those potential legal changes that seem most significant for the work at hand. These aspects were also included in the discussion with the interviewed experts (see chapter 3.3).

Pseudonymous Data

Personal data means any information related to and identified or identifiable person (Art. 2 (a) Directive 95/46/EC. A data subject is then the person to whom the data relates. 'Relate' is quite broad. The Article 29 Working Party proposes the following interpretation: "data relates to an individual if it refers to the identity, characteristics or behaviour of an individual or if such information is used to determine or influence the way in which that person is treated or evaluated." (2007, p.15)

The introduction of the new concept of pseudonymous data is highly important. It introduces a data class between personal and anonymous data. Pseudonymous information is a new type. It illustrates how data protection is becoming increasingly a risk based approach, as there is a scale: data is no longer either anonymous or identifiable, it can be in between. Experts then assess the risk of this data is endangering data protection rights.

Profiling

The legislation states specific rights and obligations in the case of profiling[17] (for targeted advertisements). Despite the broad definition of personal data, there are parts of the profiling process that are not applicable to the GDPR. A distinction between group and individual profiles is made. Former relates to a shared attribute of a group of individuals while latter refers to an attribute of one, singled out individual. The new GDPR incorporates such a distinction of profiles. It does not apply to group profiles based on aggregated anonymized data if from a sufficient number of different people.

As a consequence, data mining and interpretation (i.e. profiling) are not covered, since it is assumed that anonymisation has to be performed properly beforehand. Data is, however, still linkable to a unique user through certain means reasonably available to certain actors. This is the case especially for OSNs as direct personal data out of the individual user profiles is used, and an OSN is likely to possess the means to establish a link between data and the individual user. This is, in fact, the main asset of most OSNs.

The data mining process can be summarised into eight stages (Heyman & Van Dijk, 2013), of which only step two, three, seven and eight are covered by data protection law. The first stage is preparation; specifying the problem domain and purposes of data analysis. Second, Data Collection, the stage of constructing the database to be mined. Third, Data preparation; the stage of preprocessing the data for anonymisation and elimination of noise. Fourth, Choice of means; the stage of selecting the processing algorithms and mode of representation. Fifth, Data analysis; the stage of actual data mining (i.e. profiling). Sixth,

[17] Profiling is the act of segmenting a group of data according to certain shared attributes. The attribute is often created through data mining, the process of finding new patterns or information through algorithms. For more information on data mining see (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

Interpretation; the stage of evaluating the established profiles. Seventh, Application; the stage of applying profiles to a user based on his or her data. Eight, Measures; the stage of making decisions on the basis of the individual profile application.

Steps 2, 3, 7 and 8 trigger the applicability of several transparency rights to be informed about the existence of profiling, its purposes, the categories of data used and the available rights to intervene in this process (complaint, objection, rectification, erasure). Step 8 also triggers the specific right not to be subject to measures based on profiling. Because these are measures based on profiling, a special transparency right is applicable and the data subject has to be made knowledgeable about the logic, significance and envisaged consequences of profiling, irrelevant of the question whether personal data have been used.

Self-Regulatory Tools

Self-regulation is precautionary action by data controllers[18]. The introduction of precautionary and self-regulatory tools might become obligatory to them. Such tools include data protection impact assessments, data protection by design, accountability & responsibility, data protection officers, certification, standardization or codes of conduct.

Sanctions

One of the issues with previous data protection directives was the absence of sanctions. Organisations are only persecuted if a data protection authority (DPA) launches an investigation or if data subjects file a complaint. With the proposed regulation a specific sanctions regime for infringements is introduced. Companies are eligible to a fine if they did not protect personal data within reasonable effort: according to Art. 79 a fine up to 100 Million EUR or up to 5% of the annual worldwide turnover.

DPA's would then be able to compel companies, whereas they have to rely on other courts to interpret and apply the law today. These administrative sanctions may, however, only be applied if one can prove that a data controller did not undertake reasonable data protection efforts with regard to data use. This aspect is interesting for the work at hand, as now data controllers have to take into account the risk of being caught. This might result in more compliant and proactive data protection behaviour.

Risk-based approach

In essence, these aspects all result in a more general risk-based approach to data protection. Pseudonymous data are based on the risk of identification. Profiling based on pseudonymous data does not trigger obligations or user rights. Many of the new precautionary, self-regulatory tools originate from risk-management practices. The compliance in applying risk-based tools constitutes an extenuating circumstance in determining sanctions. An issue in this regard is that the definition of data breaches, data protection and as a result also privacy to an extent are likely become expert definitions by a new industry of privacy assessing actors.

Next, we provide some insights into the current situation regarding privacy in the USA. This is necessary, as most innovation and businesses related to personal data originate there. As a consequence, through the market, the US' privacy perspective has an effect also on Europe.

[18] see section 2.3.3 below for examples

2.4. A Note on Privacy in the USA

The US approach to privacy is quite different from the EU, the Federal Trade Commission (FTC) has been guarding privacy from a notice-and-choice framework and a more recent harm-based model (Ohlhausen, 2014).

The notice-and-choice model is based on the Fair information practice principles and focuses on the following four principles: "(1) businesses should provide notice of what information they collect from consumers and how they use it; (2) consumers should be given choice about how information collected from them may be used; (3) consumers should have access to data collected about them; and (4) businesses should take reasonable steps to ensure the security of the information they collect from consumers" (FTC 2010c, p. 7 in Ohlhausen, 2014, p. 5 [emphasis in original]). The notice and choice model is similar to our data protection approach where data subjects have to give freely given consent based on information with regard to the purpose and nature of data processing operations. Data subjects should also have access to their data. The responsibility to guarantee these rights is also in the hands of the data controller, the company that solicits this information.

The harm-based model is new, here notice-and-choice is too costly for firms and therefore another principle is used: "the harm-based approach asks whether a firm's practices cause or could likely cause physical or economic harm, or "unwarranted intrusions in [consumers'] daily lives" (Muris 2001)." (Ohlhausen, 2014, p. 5) Here the FTC is no longer a regulator but an enforcer through direct actions. Actions are taken through two principles unfair or deceptive acts or practices. In order to provide clear guidance, these principles and their interpretation are explained through guides. What is most interesting is that the FTC can act regardless of the promise made to the consumer in case of unfairness and secondly, it presupposes that the market will self-regulate. It means that consumers will find other companies that guarantee a higher privacy standard.

2.5. Existing Tools and Initiatives

USEMP aims at developing a framework that will empower users by enhancing their control over the data they distribute or interact with. To reduce the existing asymmetry between data processing and control means available to users and to other actors in the ecosystem, concrete tools could help and which we are now considering next. Today, various examples of such tools can be found in the digital world (see also USEMP D4.1 and D9.3). This section presents an overview of interesting examples, to illustrate the different characteristics of tools. They differ in function, as some intend to increase the transparency of data re-use while others go further and aim at providing monetary value in exchange for such re-use. They also differ in their positioning in the value network, i.e. in which actor they affect most directly. Another important factor for differentiation in this regard is who introduces the tool: some are provided in a bottom-up fashion, but also the industry itself is promoting certain initiatives according to self-regulating principles. Most importantly, nothing exists so far that offers all functions of the envisioned USEMP tool in one tool. This is probably the case because the market to provide this is not interesting enough yet.

Transparency

Transparency is about awareness and essentially also about control. Privacy Feedback and Awareness tools, in this regard, have the basic aim to make users aware of (unknown) processes underlying information sharing, like third parties tracking users (see USEMP D4.1

for a comprehensive list). These tools are often provided as browser plug-ins. ***Ghostery***[19] is arguably the most known example. It promotes transparency through understandable visualisations of what parties are tracking the user.

Other examples are ***Privowny***[20] and ***Abine***[21], online privacy companies providing simple tools for consumers to control what of their personal information third parties can access online.

The most widely-spread plug-ins that concern privacy in a fashion are, however, probably adblockers. They allow users to prevent advertisements from being downloaded and displayed, and to prevent tracking. It is noteworthy that the prominent, free tool ***Adblock Plus***[22] is now accused of allowing paying companies such as Google, Microsoft, Amazon or ad platform Taboola to get through its filters (Cookson, 2015).

Personal Data value & Monetisation examples

Datacoup[23] is the prime example of the approach of evaluating personal data value and monetising personal data currently testing in the open market. It enables a user to connect data of different online accounts. Then, a profile is created that provides an overview of the user's data for potential data purchasers. The profile discloses the individual data points (i.e. attributes such as gender, education, monthly spending) that are received from each connected account. Each point is allocated a high, medium or low value, which contributes to the overall base price listed for the user's data. The value is determined by the current demand in the data marketplace for that data point. The price for your data is the sum of all active data points.

[19] https://www.ghostery.com/en/
[20] https://stage.privowny.com/
[21] https://abine.com/index.html
[22] https://adblockplus.org/
[23] https://datacoup.com/docs#how-it-works

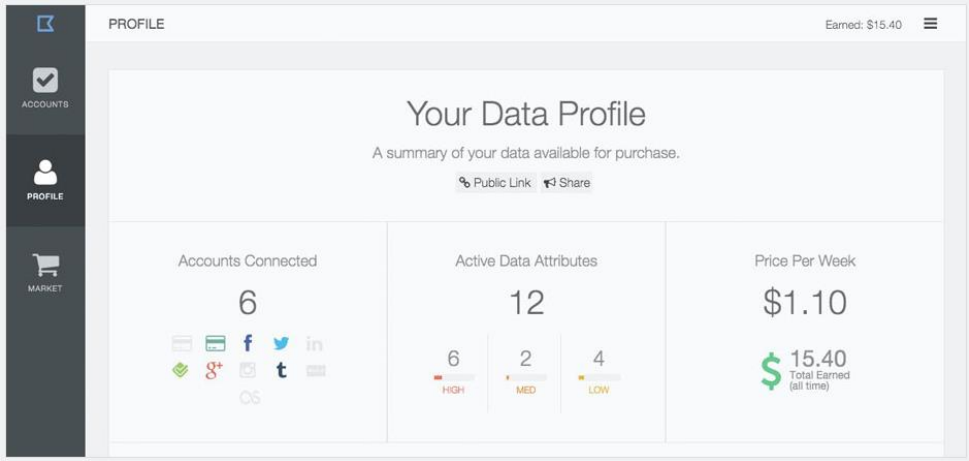

Figure 6: Datacoup's Data Profile

TheGoodData[24] follows a similar approach, but here the user lets them monetise the data and a part of this money will go to a good cause. To do so, it offers a browser extension (based on the open source software offered by Disconnect.me[25]) that prevents unwanted data tracking. The data, which the user allows to share, is then anonymously processed and traded with retargeting networks, in line with the user's preferences. The earnings are reinvested for a good cause. Interestingly, the data purchasing retargeting networks pay for user data anyway, which is, however, gathered by third party without user consent or benefit.

TheGoodData

Datarepublic[26] and *Leaflad*[27] are more examples of a similar approach to reallocating the value of personal data.

Self-regulation

Self-regulation is the process whereby companies or organisations control and police their own adherence to legal, ethical, or safety requirements. Many industrial players in the personal data market promote this form of regulation as mutually beneficial for all involved. Various initiatives and project exist. Google, for example, launched the ***Contributor***[28]***,*** a project that supposedly intends to find alternatives to advertising for funding the Internet. A user can choose a monthly contribution, which is redistributed to directly to the makers of the participating websites.

Contributor by Google

[24] https://thegooddata.org/
[25] https://disconnect.me/
[26] http://www.datarepublic.org/
[27] http://leaflad.com/
[28] https://www.google.com/contributor/welcome/

YourOnlineChoices[29] is an initiative by EDAA (European Interactive Digital Advertising Alliance), which represents the European advertising and marketing industry. It primarily tries to raise awareness about behavioural advertising, and to reduce the existing fear of users in this regard. To do so, it tries to establish clarity around practices, but it also provides an "opt-out page". Here, companies are listed that provide advertisement customised to likely interests based upon previous web browsing activity. Users can not only see which companies are gathering and using data, but also control it.

Finally, an interesting initiative that should be mentioned is **Digitrust**[30], a foundation launched Rubicon, OpenX, PubMatic and several other major industry stakeholders. Digitrust was founded as a non-profit organisation aims at standardising user-IDs across the ecosystem. A system of standardised user-IDs would make cookie synching redundant; it is first and foremost about more efficiency for the whole value network. However, almost as a side-effect, also transparency of processes would increase, which could be made beneficial also for the user.

Next, we provide a short discussion of value propositions. Value Propositions can be regarded as the incentive offered to a user to actually use a service. This is relevant, because users provide their data because they have an incentive to do so.

2.6. Value Propositions

According to Dubosson-Torbay et al. (2001), a value proposition is what a business offers as substantial value to a target customer, for which the customer is willing to pay. Put differently, it is a promise of value to be delivered and a belief of the receiver that the value will indeed be delivered and experienced. Social media and social networks offer such a proposition to businesses, being an integral and indispensible part of their marketing and communication with customers. This is the case because they facilitate the addressing of customers and the gathering of information about them.

For users of OSNs, value propositions are different and, according to statistics, they must be very good: the average global Internet user is said to spend two and a half hours on social networks daily, and Facebook for instance reportedly has 1.2 billion monthly active users. Thus, users supply a high amount of data on ONSs and indirectly to third parties. Data disclosure is a direct trade-off for the user and different incentives are presented to the user

[29] http://www.youronlinechoices.com/uk/
[30] http://www.digitru.st/

to do so. Even if they are aware of other actors using this data for other purposes, they do so because of the value proposition; because of the value they expect to experience in return.

The incentive to participate on a social network and the value proposition for participation is most of all socially driven, being part of the network to see and to be seen. In this regard, social consumption or word of mouth and brand advocacy plays a role. Other incentives can be material advantages such as monetary, discounts, special deals etc. as Foursquare/Swarm demonstrate.

The fulfilment of certain concrete needs (gratification) is another important value propositions. Google's various services offer solutions to concrete needs and in return users disclose their data, as does Waze, or PatientsLikeMe. In this case, the disclosure of personal data by users is probably unproblematic, whereas its re-use, the secondary use is possibly problematic. Put differently, the re-use of personal data can be useful for users and thus an incentive or value for them. In this light, the industry puts forward two main arguments: first, that the internet is financed by advertising and thus could not exist in its current form without the re-use of personal data; second, that advertising would disturb the digital user experience even more if personal data would not be taken into account.

3.Methodology

In this deliverable we are tasked with the development of a framework that will empower users by enhancing their control over the data they distribute or interact with so as to contribute a socio-economic perspective on value of personal data.

Yet, no systematic methodology for estimating this value can be detected. The value of personal is subjective, i.e. it only arises in relation to a specific function. The task investigates business opportunities and challenges and evaluates the question of value of personal data as well as sustainability of personal data protection more generally. In doing so, the deliverable focuses on the industry perspective. Whereas user requirements are specifically taken into account by the project, requirements of affected businesses are still missing or misunderstood.

In order to do so, the following methodology is deployed. First, to provide an overview of the status-quo, to map and describe the ecosystem, and to draw out the simplified value network, desk and scoping research is conducted. The knowledge of the status quo is necessary to provide a sound foundation for identifying the ways in which the proposed USEMP business ecosystem could evolve in the future. A suitable approach to business modelling (see Section 3.2) provides the necessary framework for the creation of value networks as well as in general for the work at hand.

Several actors within the defined value network are then interviewed providing indispensable insights. This empirical research is conducted in the form of semi-structured expert interviews (see Section 3.3). The second part of the deliverable thus elaborates the information gathered through the interviews.

The final part constructs and evaluates different potential futures of the personal data market with a focus on Europe. It is based on scenarios developed to describe potential futures and improved in an iterative fashion (see Section 3.4). As a result, most realistic outcomes can be identified in line with the needs of real-life stakeholders, prioritising economically and strategically viable scenarios based on mutual benefits.

3.1. Research Question

The state of the art presented above is reflected in the research questions we intend to answer within this deliverable. General aims are to assess the value of personal data, to analyse the value network, participating actors, their roles and relationships. We focus on the value of data and the kinds of data exchanged as commodities, but also as intangible assets. This is necessary to determine how an economically and strategically viable personal data management tool can be established.

Our hypotheses and research questions were developed partly in collaboration with the legal experts and work package partners from ICIS. This interdisciplinary starting point for our research was deemed necessary to provide consistent outcomes throughout work package 3. Based on their input, we take into account the general shift to a more risk-based approach to personal data, as promoted by the proposed General Data Protection Regulation. This

includes that there is no longer a strict distinction between anonymous and identifiable personal information (pseudonymisation[31]) in legal terms. Consequently, we take into account any data regardless of its anonymity status. The risk-based approach further entails potential sanctions for imposed for violating laws. We pay special attention to self-regulatory initiatives on the industry-side.

This leads us to our main research question guiding our work: **How can a personal data management tool be established that actually supports users in their providing of personal data, while being economically and strategically viable for businesses (i.e. sustainable)?**

Following research sub-questions need be considered in order to answer the main question:

1) **What is the value of personal data?**
    a) As there arguably is no objective value of personal data, the most interesting question is rather how (economic) value is created with data. The direct value is low for all involved parties, it only arises in relation to a specific purpose or need; thus indirect value plays a decisive role. Personal data of an individual has very little value for most other actors than the individual itself, but aggregated data, in big volume is a critical asset in the data economy.
2) **What are mutual benefits of companies and users that can arise through the USEMP tool?**
    a) Due to the status quo of the personal data (i.e. businesses' dependence on it and users' dependence on the Internet in general and free services in particular), a USEMP tool cannot aim at cutting all data flows from the user to other actors. While this interaction exists, the question is how the user can be empowered in sharing data according to his/her wishes.
3) **Where in the value network should the tool be located?**
    a) Figuratively speaking, the USEMP should be "located" in the value network where it has the most effects. In other words, a tool will affect certain actors more than others, thus it needs to affect those actors the most, which have the most impact on the whole ecosystem.

3.2. Business Models and Value Networks

For the creation of value networks, we describe the relations between actors, their roles and how data is relevant within this network. A *role* is defined as a discrete set of responsibilities, actions, and authorisations that together constitute a coherent logic for a value-adding activity. It can "potentially exist as a commercial entity in the marketplace, with its own cost and revenue balance." (Ballon, 2007, p.10). An *actor* is an entity that encapsulates a certain set of roles, even though concrete roles might be shifting. The concept of an actor still resides at a level of abstraction that not necessarily corresponds fully with current real-life organisations. Relationships are the interactions between roles or actors (usually shown as arrows in diagrams), depicting a monetary transaction, service or a tradable good (Ballon, 2007).

[31] With legislation is moving to a risk based approach, data is increasingly in a grey area, where a risk of privacy breaches and identifiability must be assessed. For anonymous data, the question is what is the risk that this will be re-identified.

The actors, the roles they play and the relationships between them constitute the building blocks for any value network. Allee (Allee, 2009, p. 6) summarises that “a value network is any set of roles and interactions in which people engage in both tangible and intangible exchanges to achieve economic or social good.” A current real-life organisation (an individual, institution, company, organisation, etc.) with an interest or stake in the outcome of a certain action is a stakeholder. All interviewees can be considered as a stakeholder or at least as belonging to a stakeholder.

To aid our work, we utilize the business-modelling framework provided by Ballon (2007), simplified in Table 2. It has been used in various analyses, in particular applied to the media and telecom industries. It consists of four abstract layers; *value network*, *technology design*, *financial model,* and *value proposition*. Each layer is built on certain integral parameters, but not all layers and parameters are of equal relevance for the work at hand.

Value Network Parameters	**Functional Architecture Parameters**	**Financial Model Parameters**	**Value Proposition Parameters**
Combination of Assets	Modularity	Cost (Sharing) Model	Positioning
Vertical Integration	Distribution of Intelligence	Revenue Model	User Involvement
Customer Ownership	Interoperability	Revenue Sharing Model	Intended Value

Table 2: Business Model Matrix (Source: Ballon, 2007, p.10)

Our task focuses on *value networks*, the layer that is most significant for evaluating the interplay of actors (here focused on user data). Its parameters deal with the architecture of actors (physical persons or corporations mobilizing tangible or intangible resources), roles (business processes fulfilled by one or more actors with according capabilities) and relationships (contractual exchanges of products, services for financial or other resources). Concretely, the parameters can be defined as follows.

The *Combination of assets* focuses on the usage and combination of resources (the most important asset in our case is data), which might lead to synergetic effects and new resource combinations. The *Level of vertical integration* relates to the scope of tasks one firm takes over in the value creation process, whereas the higher the level of ownership and control over successive stages of the value chain, the higher the vertical integration (OSNs demonstrate a high level of integration in our value network). The *Customer ownership* describes the relationship with the end customer, shaped by the access to key information of the customer, the type of contact (the actor with access to key customer information is in a good position in the personal data value network).

3.3. Expert Interviews

The empirical research is conducted through expert interviews because it “lends itself as a data generating instrument in those cases in which the research focuses on the exclusive knowledge assets of experts in the context of their (ultimate) responsibility for problem solutions.” (Pfadenhauer, 2009, p.84). Although we are less interested in the experts’ responsibility for problem solutions, we are in need of their expert knowledge because this is not codified sufficiently for the personal data market. As previously mentioned we wish to

have an overview of the value network that uses or gathers personal information in online and social media campaigns. We want to know what the required roles are, which actors occupy these roles, and how they interact. An overview of the value network will help us identify how a tool for personal data management can be most effective.

The interview method allows us to gain a deeper understanding of meaningful themes, practices, and relationships from the interviewees' own perspectives through the collection of qualitative user-related information. It allows for a richer means to learn about possible impacts, efficiency, barriers about issues related to personal data in a commercial context. However, a volunteer bias factor of recruitment of interviewees can be a drawback as it is likely that particular types of users donate their knowledge and time to this project, which may be reflected in a somewhat idiosyncratic outcome of the analysis. With such limitations in mind, however, the mixture of methods (see 3.4), and existing research (SOTA) guiding the empirical framework presented here, are addressing these possible limitations.

3.3.1. Interviews

We conducted a series of interviews (N=9) with diverse actors from the ecosystem, to get first hand insights. The gathered data is analysed in the form of a thematic analysis. This approach promises structuring and rich description of our heterogeneous data set. It goes beyond simply counting phrases or words in a text, identifying implicit and explicit ideas within the data. The themes become the categories for analysis (Guest, 2012). The steps of the analysis are; familiarization with data, searching for themes in data, reviewing themes, defining and naming themes, and producing the final report.

The set-up for the interviews is the following: The interviews last approximately one hour. The interviews are semi-structured and are conducted following an interview guide. Questions follow a similar structure and validation procedure as the other methods used in this project. Also, several informants are asked to review, comment on and check the drafts they contribute to, thereby contributing to the validity. The interviews started with a general introduction with regard to the USEMP project and the questions we wanted to ask them during the interview. An elaborate list of questions can be found below in

ANNEX A) Interview Questions. The interviews do not always run chronologically (i.e. semi-structured interviews); experts sometimes wish to comment more on one topic or started of with the topic they found most important. We emphasise the following elements:

A. Situate your company or sector
B. Situate data use within your company or sector
C. Situate the current and future impact of the General data protection framework
D. Where would you put user-centric data management tools?

If the expert allows it, the interview is audio recorded and transcribed. The contributions are codified and thematically analysed. We inform experts about the deliverable and that we will invite them to comment on it before we draft the final version. This ensures trust between the interviewer and the expert but it also adds to the proposed iterative approach. We report each expert by his or her name and organisation. (A complete list can be found in ANNEX B) Contacted experts and elite.)

3.3.2. Expert selection

The expert selection and interview-guide content is motivated by a distinction between experts and elites. **Experts** offer broad and general knowledge about their sector or associated value network. **Elites** may also be able to offer this, but they are actors of the value network and are asked to situate their company as an active part of this value network. This distinction is arbitrary and based on two disciplines that utilise expert interviews for their research, with two distinct definitions: Pfadenhauer (Pfadenhauer, 2009) situates 'experts' in the qualitative interview methodology, and international business studies refer to 'elites' (Bandara, Indulska, Chong, & Sadiq, 2007; Welch, Marschan-Piekkari, Penttinen, & Tahvanainen, 2002). Both terms were kept as they point out different aspects important to our sampling strategy.

The **'expert'** is conceptualised as a meta-specialist, who "knows what the (respective) specialists know in their fields of knowledge - and how what they know relates to each other" (Pfadenhauer, 2009). The expert is thus an observer or even an outsider who does not necessarily partake as an actor within the value network. These experts can also be referred to as **observing experts**[32]. **'Elites'** are defined as: "an informant {…} who occupies a senior or middle management position; has functional responsibility in an area which enjoys high status in accordance with corporate values; has considerable industry experience and frequently also long tenure with the company; possesses a broad network of personal relationships; and has considerable international exposure." (Welch et al., 2002, p. 613) As they use their knowledge, they reshape their environment: "By becoming practically relevant, the experts' knowledge structures the practical conditions of other actors in their professional field in a substantial way." (Bandara et al., 2007, p. 166) We will refer to this kind of expert as **participating expert**. Although elite and expert were used, they both referred to a professional who does not only have expert knowledge but is also participating within his field of expert knowledge.

[32] Note that no interviewed expert was completely outside of the analysed value network. In this sense, the division is strictly analytical. Sector organisations can influence the value network through lobbying, representation, providing benchmarks, best practices and lawsuits on behalf of a group of actors.

Participating experts provide insight as an actor inhibiting the value network; they are selected for their position. Observing experts are asked to situate the whole rather than one of its parts. Both experts will explain how the value network is structured.

3.4. Scenario Development and Workshop Iterations

Rather than an attempt to predict the future, a scenario is a simulation of certain possibilities. Plausible outcomes are established; preferred visions of partners can then be evaluated towards real business decision-making processes. Scenarios developed for this deliverable are of exploratory nature, their setting-ups founded on aspects of the existing ecosystem. Its implications evaluated. Factors necessary to consider in course of the research are identified in accordance with the business-modelling framework and its parameters.

For USEMP, the evaluation of scenarios was meant to be undertaken by means of SimBU (or its successor BEMES) in a workshop setting: “a graphical tool to business experts for creating and evaluating high-level business models.” It allows “to get an estimation of the profitability of a certain model and define value and money streams quickly and transparently.”[33] This tool strongly depends on the financial parameters of the business model framework, and only provides workable outputs with the input of exhaustive financial information. Due to the innovative and exploratory nature of the USEMP tool, such information is not available. Furthermore, at this stage of the project and the field it is located in, we regard it as necessary to look beyond relations created by money and incorporate both tangible and intangible assets, such as trust and reputation created through transparency (tools).

As a result, we needed to rely on another method to evaluate and valorise our research. We developed scenarios that describe potential futures. The scenarios are developed against the backdrop of USEMP’s initial scenarios, as developed and described in D2.1. They are improved in an iterative fashion. Stakeholders provided feedback on decisive aspects of the suggested scenarios. They explained what they expect to happen in the future, what they expect from other actors or from new roles. Among the most important aspects in this regard was the implementation and use of a transparency tool. The iteration was conducted through several interviews and additional contacts at different stages of the research.

[33] https://forge.fi-ware.org/plugins/mediawiki/wiki/fiware/index.php/Business_Modeler_-_User_and_Programmer_Guide

4.Results of Empirical Research

4.1. Ecosystem Description Based on Interviews

“Figure 1: The Value Network of Online Advertising”, as introduced in Chapter 2, simplifies the personal data market in relation to digital marketing and advertising. As described above, the marketplace is bipolar, advertisers and publishers being the primary actors at the sides, with those in between facilitating the interactions of the sides. The user is somewhat outside the network, despite its essential role of consumer and data-producer. Developments of recent years have also shown the special position of OSNs in the ecosystem. Their evolving roles are a major concern of other, related actors, as our interviews have illustrated.

In order to get first hand insights, we conducted a series of interviews (N=9) with diverse actors from the ecosystem. We managed to recruit experts and elites distributed well over the value network. It is important to note, however, that experts like Kimon Zorbas, Ionel Naftanaila or Chris Payne have first hand knowledge of the ecosystem (and provided essential insights), but do not perform any value-related roles per se. They all work for organisations that represent broader interest groups. These parties are defined as an observing expert.

The following report of our empirical research describes the ecosystem in-depth, based on the interviews. It commences with some general thoughts on data and value in the value network, before it goes into detail for each actor or grouping of actors as determined above in chapter 2. Chapter 4.1.1 and 4.1.2 discuss the interviewees opinions on user-centric data management and the GDPR respectively. Chapter 4.1.3 of the report then delves deeper into the concrete use of data in the ecosystem.

4.1.1. Value Network: Actors, Roles, Relations

The ecosystem revolves around data in the wider sense. Data is its most valuable asset, and involved companies are more or less data-driven. Data, i.e. not necessarily personal data but audience data, is used by advertising as well as e-commerce, logistics and more; to improve a product or service, to improve the marketing of products or services, to gain insights and to sell such insights. Especially for marketing, audience data is essential, due to personalisation; place, time, context.

With automation and connected systems, the utilisation of data became possible at scale. In general terms, data drives decisions, and, according to Joelle Frijters[34], the industry today could not anymore function without it. Thus, data-driven advertising allows to find the right audience and interact with it appropriately. In fact, the collection of data occurs by default, even if the collecting actor might not have any immediate use for it. Secondary use of data (e.g. selling of data) motivates this collection. In this case, the purpose specification of data collection is problematical.

[34] Joelle Frijters interviewed by Jonas Breuer, 14/01/2015, Antwerp. Joelle is co-founder and CEO of ImproveDigital, a European provider independent publisher monetisation technology. (improvedigital.com)

Still, several of the interviewees agreed that it is hard to pin down data in marketing operations, as it is pervasive. In terms of value, it is also difficult to quantify, although there can be direct values derived from specific purposes. Also, each actor collates its own value in using data. In other words, data has no intrinsic value (in particular not personal data of a unique user, as stated clearly by several interviewees, e.g. Niels Baarsma[35]). Facebook illustrates this point, as it has very precise data with its enormous database, but is still making relatively little money with it. Conversely, according to Mario Vanlommel of BeMobile,[36] the collection and sharing of data for only one specific purpose would entail "gigantic" marginal costs.

Social media and social networks are critical in this regard, according to Lien Brusselmans[37]. This is also a reason for the success of the social media management tool Engagor, which offers engagement, monitoring and analysis options to large companies. As Chris Payne[38] pointed out, social media has become the first channel of interactions and communications between companies and customers. It is most significant because advertisers need the first party information that arises from such direct interaction. As a consequence, social and digital cannot be separated in terms of marketing, and social media is more than just a marketing device.

Regarding the interaction with customers, the term vendor relationship management (VRM) is interesting. It describes how customers can become more in control (in contrast to customer relationship management) through appropriate software tools. Privowny (see chapter 2.5 above) can be seen as such a tool, offered to companies to make their data uses more transparent. As a consequence, users' might trust more in these companies and thus choose them over others. Kimon stated, he thinks that companies are actually seeing the value of such tools at least in the mid- to long-term, although there might not be an immediate ROI.

How is data made valuable in the ecosystem

The interviews have made clear that, without certain kinds of user information, the industry could not function as it does, and users' online experience would downgrade considerably; in particular information collated via cookies. Furthermore, it is important to note that Chris and Joelle stressed that neither advertisers nor publishers have an interest in distracting the online user experience too much through marketing/advertising. These actors interact directly with the user, and depend on the user.

The data streams through the value network and touches many different actors. As described in Chapter 2, additional actors have complemented the value network with new roles over the years, and add value to the traditional chain of advertising. Thus, also value

[35] Niels Baarsma interviewed by Jonas Breuer, 23/01/2015, Brussels. Niels is co-founder and CTO of Yieldr, a demand side platform provider.
[36] Mario Vanlommel interviewed by Rob Heyman, 30/01/2015, Brussels. Mario is Technical Sales Engineer at Be-Mobile, a leading provider of traffic and mobility content and services for the automotive industry, mobile, media and government road operators.
[37] Lien Brusselmans interviewed by Rob Heman, 21/12/2014, Gent. Lien is Marketing & Communication Manager at Engagor.
[38] Chris Payne interviewed by Jonas Breuer, 21/01/2015, Brussels. Chris is Public Affairs Manager at World Federation of Advertisers. The federation is a global organization representing marketers and advertisers. (http://www.wfanet.org/en)

moves in complex ways. Actors add value to the data or derive value from it, and it has different kinds of value for the different actors, and different kinds of data offer value to the different parties.[39] However, they all share the overall derived value from the use of data. Interestingly, the actors constituting the midfield of the value network (the intermediaries) have no direct contact with users, but have most touch points with their data. The use of data is their core business, and privacy regulators are concerned about their operations. However, in their role of intermediary they argue that it is hard to change the system.

In this context, Kimon Zorbas added, "when you have a company acquiring another company in the digital world, it is for the data."[40] Therefore, the idea exists to establish a standard for accounting for the value of data as an asset, because no clear rules exist. Despite being an asset, it is difficult to say what exactly is valuable about a database. This points back to Facebook's difficulties to monetise its user- and data-base.

Another issue in this regard is that the quality of data in the current internet marketing ecosystem seems to be very low. Thus, tools that enhance transparency could also raise and clarify the value of data for the industry. Visualising data points and streams (as some existing tools do) are helpful in this regard. In this light, Niels Baarsma[41] also advocated the idea to make cookie synching more transparent and effective, as advantageous for both the user and the industry. This again points back to the concept of vendor relationship management (VRM). It could increase trust in an ecosystem, which is lacks trust within as well as from the outside today.

Advertisers

Theodores Michalareas[42] of Velti stressed the importance of distinguishing advertising and marketing. According to him, advertising is the act of buying and showing an ad, and the value of data mainly lies in paying less due to less 'waste' of ads (thus used for evaluation; predicting or describing ROI). For marketing, then, data is used to better understand what a customer wants and the context of the customer; thus how, when and where to reach him/her (profiling and targeting).

As a consequence one can argue that targeting and other data-driven processes offer more potential for marketing than advertising.[43] Here, it is used to assess all that a user or customer does; churn, where they are in the marketing funnel; what their customer lifetime value is. The main insights about this perspective were gained in the interview with Chris Payne of the World Federation of Advertisers (WFA), which presents the interests of both advertisers and marketers. Therefore, it cannot be distinguished rigorously when reporting on the empirical studies.

He explained that there is a trend for brands to own the data, due to the issue of data-quality (i.e. veracity), but also security concerns. For them, the first party information that arises from direct interaction (in particular via social media) is at least as valuable as data bought from

[39] Chris Payne (21/01/2015, Brussels).
[40] KImon Zorbas & Ionel Naftanaila interviewed by Rob Heyman, 28/01/2015, Brussels. Kimon is Director at the Digital Business Consultancy Group (www.dbcg.eu).
[41] Niels Baarsma (23/01/2015, Brussels).
[42] Theodores Michalareas interviewed by Rob Heyman, 19/12/2014, via Skype. Theodoros works for Velti, provider of mobile marketing and advertising technology.
[43] Theodores Michalareas (19/12/2014).

other parties. Data-related operations generally seem to be preferred to be in-house.[44] Vertical integration of roles is interesting, especially because the tracking across screens/devices is the biggest challenge at the moment. Niels Baarsma confirms.[45] Ronald Siebelink[46] also pointed out that data of past marketing campaigns is becoming an asset to advertisers. Before the advent of DSPs such as Rocket Fuel, the data was kept within the advertising agency and if the advertiser changed agencies, the data was lost. This loss of data is seen more and more as a switching cost and what is more, the advertiser has paid for this data so it should be his. Lastly, the data of past campaigns can be used as input for machine learning algorithms to optimise future campaigns.

Advertising works well enough with pseudonymised data, which is enough for segmentation. Direct marketing, which is based on personal data, is much less common, and targeted advertising rarely relies on personal data. The data is used for the following purposes: targeting segments, attribution and feedback. In the first case, users have to be grouped in segments that share the same attributes. With attribution, users are tracked and have to be uniquely identified to report that there exists one user who viewed or interacted with a particular advertisement. Lastly, users are tracked to provide feedback to report on the advertising or marketing campaign. In this case, users are again segmented in categories corresponding to the marketing funnel. Awareness, engagement, action and advocacy...

The collection of such data and its holistic use does not pose high risks for consumers. The data, which is used, is necessary for advertisers to be relevant with their ads: tailored communication, which should also be engaging or interactive. According to Chris Payne of the WFA, the combination of UGC with commercial data plays an important role here.

Publishers

Joelle Frijters confirms that “today most data that is used is owned by advertisers”, collected by advertising agencies or the advertiser itself. The data is mostly cookie based. She continues, stating that “an ad is worth much more if you know something about the audience”, in combination with the content of the ad.[47] In general, however, the value of ad space has constantly been decreasing, and even with attached data it is worth much less than it used to be. This is a main challenge for publishers, as they are financed to a big share by the advertising machine.

This is the case mainly because of the competition in regards of displaying advertisements. The digital world offers huge amounts of ad space, instead of only newspapers and other traditional publishers. Thus, “there is an abundance of advertising inventory”[48], which undermines the operations of publishers. Despite the competition from social media such as Facebook, publishers are highly dependent on OSNs and “likes”. A major share of their traffic is coming in social media. Thus, they cannot do anything against this leakage of value, because they depend on the traffic. In addition, the monetisation through subscriptions has constantly been dropping, as a result of a lot of free content online.

[44] Chris Payne (21/01/2015, Brussels).
[45] Niels Baarsma (23/01/2015, Brussels).
[46] Roland Siebelink provided his insights during an event, 17/12/2014, at iMinds-Smit-VUB in Brussels. Roland is head of quality, productivity and best practice of Rocket Fuel.
[47] Joelle Frijters (14/01/2015, Antwerp).
[48] Joelle Frijters (14/01/2015, Antwerp).

Intermediary: Supply Side Platform

ImproveDigital - represented by co-founder and CEO Joelle Frijters - is a supply side platform (SSP). It delivers services to the publisher and charges for the facilitating of impressions from publisher to auction. During the interview, Joelle stressed that the company's main selling point is to provide independent technology to (big) publishers in order for them to be able to stay independent. It provides real-time advertising technology to publishers who want to build their own ad ecosystem. The company deliberately decided to stick exclusively to the publisher side.

Big players such as Google, Facebook, Adobe or Microsoft own an ad ecosystem which is then offered to a publisher as a service. Google's ad server, for instance, has a market share of 80% (according to Joelle), and publishers might choose the leading player as service provider over their independence. In general she stressed how increasingly difficult it becomes for smaller players to compete in a market that moves towards consolidation and ultimately oligopoly (Google and Facebook are the examples mentioned most by her, but also by other interviewees). This independence is decisive for publishers in today's ecosystem, as the other intermediaries exist to minimise ad spending on the advertiser side, which results in lower income for publishers.

It needs to be mentioned that, due to ImproveDigital's offering of a complete ad ecosystem, their interaction somewhat fall out of the highly simplified categorisation in the value network in Figure 1. They can be categorised as SSP, but they provide a wider range of tools for big publishers to operate their own ad ecosystem independently. The technology platform they offer combines all tools necessary to do so.

According to the interviewee, ImproveDigital and its platform should be regarded as an "optimiser". Any actor buying ad inventory from a publisher has to go through the platform. "A publisher uses us to sell inventory, to Google, Facebook, or the twenty or so other agencies and buyers". To clarify, ImproveDigital is not an ad exchange, as they do not sell ad space themselves to advertisers, but only offer the technology to publishers to do it themselves.

Intermediary: Demand Side Platform

A demand side platform supplies services to the advertiser side; an agency or directly an advertiser. Yieldr is a successful company in this field, co-founder and CTO Niels Baarsma offered his insights during an interview. It offers a technology platform facilitating efficient advertising, enabling advertisers to do goal-based advertising: based on ROI targets, conversion targets, click targets etc. The advertiser can setup the platform itself (similar to to google adwords) or let Yieldr do it, according to their business rules and goals: budget, campaign liftetime, targets such as amount of conversions. Goals are measured through beacons throughout the sales funnel. The ultimate goal is, obviously, to make the customer reach the end of the sales funnel; the purchase. According to Niels, the average worth of this can be said to account for five euros. This process takes place in real-time**.**

The operations of a DSP, and their use of certain kinds of data, are founded on two main pillars: prediction and conversion. Predictions of impressions, clicks, and conversions are calculated based on extensive data points. If the prediction for a bid is sufficient to clear the floor price set by the SSP, a bid is sent and the auction determines which bidder wins. The winner can then send the ad to be shown on the ad space as paid for. What Yieldr offers is the necessary bidding technology. Advertisers pay to actually participate in an auction of ad inventory.

A DSP combines information from both sides, advertisers and publishers, and facilitates interactions between them. Utilised data points also contain information about the user, such as browser version, OS, device type, location if app requests GPS access. Yieldr adds some data points, as they check how often a unique userID has seen a particular ad and how much time was in between. This affects predictions; if a user saw an ad many times and did not respond than prediction should be lowered.

As this concerns billions of requests in real-time, the technology of a DSP has to make sure that bids are only placed on relevant impressions. Otherwise, the boundaries set through the advertiser's campaign would be exceeded. Niels summarised this as giving "the system enough tries/successes in order to calculate a probability to success."[49]

Alongside prediction, conversion is most important. First, conversions are measured: an add is shown to a user, who has a unique user-ID or is assigned one then. Such IDs are based on cookies, stored in a fast memory database and exist for fourteen days on average. When this user then accesses the page of the advertiser, where a beacon is placed. The beacon calls the URL of the user-ID, which is then cross-referenced in the database, identifying the user and determining the conversion. Beacons are usually placed on multiple points on the advertiser's page. Each beacon represents a point in the sales funnel and thus x amount of value.

When this process takes place - a user with an ID clicked on an ad, went to advertisers site, purchased - each actor has to prove to the paying actor (advertiser) what part of the process they caused. Therefore, the advertiser sends an order-ID for a conversion, based on which actors such as Yieldr can prove the number of conversions they facilitated. This is then the second conversion-related aspect: conversion attribution. Based on this, the advertiser calculates the ROI.

Conversion attribution is a decisive aspect in the debate around personal data, and regulators see it very critical. On the one hand, following the user through the sales funnels by means of a user ID is necessary for the functioning of the business as it functions today. Advertisers partner with multiple different companies, all fighting for the sales, and need to determine who caused a conversion. On the other hand, a user-ID can make a user identifiable, as the advertiser could theoretically track the order-ID to a person. According to Niels, "this is an issue, but it is not restricted to display advertising. It is what drives the Internet."[50] The process is documented in the RTB (real-time bidding) protocol.[51]

Intermediaries: Ad aggregation and Sale

As mentioned above, both ad networks and ad exchanges need to be categorised under this group, although former have been all but replaced by latter due to the advantages, which automation offers. The exchanges between buy- and sell-side are central to the whole ecosystem, because they facilitate the interaction between the sides, in the form of a marketplace. Due to this position, ad exchanges can take large shares from the traded revenues. According to Joelle, possibly up to 40%.[52] This is also the reason why most actors,

[49] Niels Baarsma (23/01/2015, Brussels).
[50] Niels Baarsma (23/01/2015, Amsterdam).
[51] http://www.iab.net/guidelines/rtbproject
[52] Joelle Frijters (14/01/2015, Antwerp).

in particular bigger ones such as Google (as Doubleclick), Yahoo (as RightMedia) and Facebook want to establish a strong position in this group.

Information Broker

Although no information or data broker, or similar actor, was interviewed directly for this study[53], a significant aspect should be notified in this regard. Several of the interviewees have a relationship with this position in the value network. Interacting with data brokers is often about outsourcing the responsibility for the data, in terms of privacy, but also veracity and security. Mario Vanlommel of BeMobile[54], for instance, stated that, due to data minimisation, the company purchases external data: then the supplier carries any risk and takes care of consent and privacy.

This is interesting for the functioning of the whole ecosystem. One can argue that actors prefer to bypass issues that arise with the use of data by externalising the risk. In this way, they can act as if the data was neutral. This is arguably related to the low trust, between actors in the value network as well as from the outside towards the whole ecosystem of display advertising. Actors such as Engagor and Rocket Fuel have many different clients of whom many are competitors. Thus, they need to appear as strictly neutral service providers, or separate clearly their interactions with the different clients.

OSN

It was already mentioned before, that OSNs occupy a particular and powerful position in the ecosystem. The reason for this is most of all their unmitigated contact to the user. As Joelle puts it, “real problem are these big companies”, regarding other actors in the industry that feel threatened, but also for the protection of personal data. For a consumer, it has become increasingly difficult not to use them. As Kimon Zorbas clarifies, “Yahoo, Facebook or Google are [...] companies that have a vertical and a horizontal model to combine the data.”[55] This pervasiveness, coupled with non-transparent practices, enables to do anything with users’ data.

Knowing more about users than other companies is the biggest competitive advantage. Selling information about users is a successful business model, practiced notably by Facebook and Twitter. However, as the interview with Joost Roelandts[56] confirmed, it is not the only model for an OSN. Twoo’s revenue, for instance, only consists of advertising for a minor share, the major share being generated by its freemium model.

As mentioned before, publishers have specific problems with OSNs such as Facebook or Google. As a result of, for example, Google’s expansion, they interact directly with publishers, but are also competition as they offer the same advertising space to advertisers. Other intermediaries in the value network exist specifically to avoid such conflicts of interest. Advertisers and brands, on the other hand, might generally worry about monopolistic or

[53] Unfortunately, no representative of this actor was available or willing to provide insights for our research.
[54] Mario Vanlommel (30/01/2015, Brussels).
[55] KImon Zorbas & Ionel Naftanaila (28/01/2015, Brussels).
[56] Joost Roelandts interviewed by Rob Heyman, 06/02/1015), Brussels. Joost is COO at the social network Twoo.

oligopolistic tendencies. However, they do not have the same issue, as for them it only matters where the customers are.

4.1.2. User-Centric Data Management

Interviewees were asked about their opinions concerning user-centric data management, empowering data subjects with regards to the re-use of their personal data. In general, they do not see any notable economic potential arising from a shift of control over data management to end users. This, however, needs to be treated with care as all interviewed actors are part of the advertising value network, which is founded the re-use of user data. Several reasons could be retrieved.

First, the value network is still based mainly on cookies (although this is in transition due to the increase of mobile devices). Several actors launch their own cookies to collect the data of users online and all these cookies are different. Software that could track all utilised cookies of all actors would have to be capable of interpreting the different algorithms. This is difficult and presents already a barrier to establish user-centric data management.

Additionally, it would cost a lot of effort for a company to make data accessible to users, as Joost Roelandts[57] of the social network Twoo explained. The main problem here is that platforms are in constant evolution, changing almost on a daily basis. Thus, the scripts to offer the data to a user have to be rewritten and supported on a daily basis, too. Furthermore, the data they have is so big in volume that it would be very time consuming to find all data of a unique user. Such databases are created to store big amounts of data, to aggregate and assess it in this form, not to retrieve personal records. This concerns usage data, mainly on the back-end, which is implicit data (logging).

There is a major and significant difference to profile data; this is volunteered data, part of the front-end. This explicit data is personal. It is thus also much easier to assess and delete, contrary to implicit data in the back end. The latter is pseudonymous data, not being based on a real name but a user-ID to signify a unique record in the database. Only once has a user requested Twoo to provide all data (both implicit and explicit), which is made possible through legislation.

Joelle Frijters discussed the idea of a ‘right to pay’ for specific online services, which could forestall any re-use of personal data. This could be an option, but there are several points that render it difficult to realize. First, a right to pay would clearly favour more wealthy users, while those who cannot pay would be discriminated. Second, most people do not seem as if they would like to pay for services (as the low rate of subscription for newspapers might demonstrate) and actually accept advertising as the price to pay.

Chris Payne adds in this regard, that the biggest added-value a user can derive through user-centric mechanisms is improved quality of marketing, i.e. a better user experience: ads would be more relevant, they could be less disturbing etc. This is also in the interest of advertisers. Kimon and Ionel discuss in this light the issue of low data quality and trust in the value network: “the business view is that you have a lot of data that is inaccurate. Third party

[57] Joost Roelandts (06/02/2015, Gent).

data is often very inaccurate"[58]. Which is why user-centric data management tools can help to increase quality.

Personal data value & Monetisation

The discussion about user-centric data management often includes monetisation of personal data for the user. The non-transparent economy, which uses personal data as a currency, is for a major part about marketing. This is the reason for focussing on the ecosystem of display advertising in this work. The interviewees were asked about business opportunities and challenges posed when users monetise their personal data.

A direct monetisation of personal data, as in "I give you my data, you give me some money back" is regarded as least feasible. The main reason for this is that there is not enough value in unique users. The effort of establishing such a system thus would not be proportional to its output. Niels and Chris explained that advertisers can not pay users for their data, due to the low value of individual data points. The low CTR is the main reason for this. Programmatic buying and relevant hardware only make sense at scale, and advertisers usually pay per thousand impressions (around one euro according to Niels). Thus, pseudonymous data that can be used for segmentation is sufficient. As a result, the utilised data is not only too big to do anything personal with it, but the individual is also not interesting enough from an economic standpoint.

The value that marketers derive from an individual's information is indeed less than many think; there is a strong imbalance between the perceived value and real value. Therefore, expectations need to be managed, of users as well as of regulators. There clearly is a trade-off between providing data and receiving value in return, but this transaction between the advertiser and consumer is not direct. Several initiatives have investigated this aspect. Google Contributor (as described above) is an example, as is Bing's idea to donate for a good cause based on conducted searches. Some publishers have offered content in return for data, as well as for payments, and have redistributed value in this fashion.

Most significant in this regard is, however, the cross-subsidization of free services online. Such services are enabled through advertising, with few exceptions. The funding of free services through ads is the most significant value the user receives in return for his/her personal data. Here, the surplus that a single user derives through using a service without paying for it is higher than the value for an advertiser receives for the individual user's data.

4.1.3. GDPR

Aspects of the general data protection regulation, as it is being discussed at the moment, underlie this work and partially guides also the experts interviews undertaken. Therefore, certain criticisms and suggestions of the interviewees concerning the regulation need to be taken into account.

Most of the interviewees displayed a critical standpoint towards the regulation. It was badly drafted, contradictory, according to Kimon even "a disaster". Provided reasons included: First, too many interests and views need to be combined. Second, it aims too much at big companies such as Facebook, and as a result the needs and requirements of smaller

[58] KImon Zorbas & Ionel Naftanaila (28/01/2015, Brussels).

European firms can be overlooked.[59] It was also mentioned that a large part of the regulation is based on fear, stemming from the economy's bad reputation, and a lack of knowledge.[60] In this regard, the intention of the regulation is to punish the moguls but in reality, smaller European companies are punished. This is the case because they are easier to prosecute due to their location but also because their legal departments are often smaller. As a result the European market is weakened, which strengthens the position of international moguls such as Facebook and Google.

The European industry particularly fears that the strive for privacy and data protection in Europe will handicap their competitiveness towards other international players and their innovativeness. "The balance between economic benefits for Europe and privacy is really important."[61] An issue in this regard is that a regulation can only be truly effective it is also affects actors such as Google and Facebook, which are among the strongest agents of the personal data economy. It was mentioned that regulation should take into account company size, in order for young or small companies have enough room to compete with innovation from the USA.

Lack of clarity in existing regulation is another issue for European companies, as it is disincentive for them to invest in following the law. It is cumbersome for companies to meet existing requirements, as processes are too unclear, in terms of timeframes, certainties, transparencies, and guarantees when meeting the conditions.[62] That privacy legislation is national also presents a difficulty for European companies, as they have to adapt their products and services to each country.[63] Also in terms of definitions, new regulation must improve. It is not clear what exactly counts as personal data. Also the move to pseudonymous data was criticised, as it is unique and thus ultimately identifiable.

The interviews illustrate how the different interests and requirements of different actors affect their position towards the regulation. For example, brands and advertisers argue that the current consent regime disturbs the user experience too much, and they see that users also do not want to give their consent all the time. Also, the risk-based and proportionate approach is regarded as most appropriate (demonstrated by the USA), as well as self-regulation and clear rewards in return. On the other hand, it can be argued that risk-based is invented by the industry, as it is particularly convenient for them, but too difficult to assess and determine.

4.1.4. Data

The value of personal data or a category of personal data is relative to the goal the data is used for. Because there are many different uses and users, this value is relative. In order to understand the value of personal information we describe the different uses for data, the collection process and expectations of these different users.

[59] Kimon Zorbas & Ionel Naftanaila (28/01/2015, Brussels).
[60] Niels Baarsma (23/01/2015, Amsterdam).
[61] Joelle Frijters (14/01/2015, Antwerp).
[62] Chris Payne (21/01/2015, Brussels).
[63] Joost Roelandts (06/02/2015, Gent).

Data gathering and sharing

In general, data is gathered via unique identifiers. Cookies are the most known example therefor. These unique identifiers are associated with other attributes of a user. All parties who have no need to keep personal information will not attach personal information to these identifiers. Unique identifiers are compared for each party and common cookies or common records receive new attributes describing the cookie or the cookie owner.

Each actor launches its own cookie or other unique ID, gathering commercially relevant information. The actor has its own database with such identifiers. However, in order to interact, specifically for the real-time bidding process, the different IDs have to be matched across all actors. A mapping table, therefore, functions to map these different identifiers from different actors, and it needs to be kept up to date. This process is called cookie synching. If cookies are not synched up, a user cannot be identified and thus the bidding cannot function.

The process is time-consuming and information has to pass many partners before it can be used. It is very inefficient, as each actor needs to maintain these tables, which entails costs. It is also non-transparent, for the user, but also for the publisher and the advertiser. A side effect is that this makes it very difficult to track consent given by a user.

Hashing is important in this context. It describes the way non-identifiable information is shared without revealing identities. It is similar to cookies, one non-personal shared attribute is used to transfer new information. A hashing algorithm is created for a data set. These algorithms change personal information which two databases share, into a nonsensical code. The two lists of hashed personal information are then compared. Matches receive new attributes and non-matches do not.

Directly usable data

Unique identifiers. The most common collected information is a unique identifier, which uniquely identifies a user. This piece of information only conveys a browser or device and is unique from other visitors of the website or application. This is usually done by storing a unique feature of a user or tagging a user with a unique identifier. The most known technology to do this, is the html-cookie. The unique identifier is used to log how many unique visitors visit, but also to attribute information to this unique identifier. For example has this unique visitor already seen this ad? This has been called frequency capping and advertisers define how many times an individual should be exposed to an ad.

Verification and registration with volunteered data. The interviewed experts have not mentioned verification and registration often. Login or registration forms gather volunteered information from users. This is primarily used for users to enjoy the service and for publishers to personalise the service. Ionel[64] did point out that these particular moments, when users log in or sign up, are used to redeploy cookies that may have been deleted. So the use of volunteered information is to verify a user's ID and to tag this user as such through cookies.

Observed data. Observed information refers to data that were not asked explicitly from the user. Browsers and mobile devices communicate technical information such as screen resolution, operating system, language, etc. This information can be valuable as such or it

[64] Ionel Naftanaila (28/01/2015, Brussels).

can be used to uniquely identify users or their devices indirectly. The Electronic Frontier Foundation illustrated this point through panopticlick.eff.org, a website that tests how unique a browser is. The originally intended use for this information is to inform the server of compatibility issues, for example should the lay-out be optimised for large or small resolution.

Contextual data. Data provided by the publisher in the Real Time Bidding (RTB)-protocol. This data entails the format of the advertisement, a category to define the website and observed or volunteered data about the user who generated the impression.

First Party data. Data gathered through the first party, i.e. the user, in the form of volunteered data. It can be logged, through cookies or similar technologies or solicited via forms that are part from registration forms or required to provide a certain service or good. First party data is also gathered through direct communication on social media. It can be integrated directly in marketing campaigns or indirectly by selling it. This is for example the model of OkCupid, or better said its parent company IAC/InterActiveCorp (Downey, 2012).

Inferred. Audience data, ad exchanges monitor the behaviour of visitors across their bought ad inventory. Online behaviour is then linked to panels and from this their expected behaviour is extrapolated. Panel members with similar behaviour were more likely to buy product X or Y or click on ad U or V. RTB platforms are data-driven, i.e. they predict the most cost-efficient ad inventory and learn from past interactions. The data required as input is based on other people's past behaviour, third party data and data from advertisers.

Indirectly usable data

The data described below is only usable in conjunction with another goal or data sources.

Cost saving. The use of data saves costs for advertisers or increases profits for publishers because the data more efficiently increases ROI for advertisers. Profit increases for publishers if advertisers have a guaranteed ROI through particular publisher inventory. Here, the data driven platforms (such as DSPs and SSPs) function to predict which inventory this would be.

Targeting data. Targeting data are used to include or exclude segments from an advertising campaign. The data can come from the advertiser, a data broker or the publisher. Here, the value of the data lie in their reliability; are the segments corresponding to the actual target group? Trust in third party data is low because the source of this data is not transparent. Kimon Zorbas and Ionel Naftanaila[65] referred to a tactic of Nielsen to address this issue. Her Nielsen compared the reached audience with the intended audience and Facebook's data on the reached audience. By attaching Facebook data, which is assumed to be reliable, advertisers were able to see the difference. Otherwise advertisers only receive reports from advertising agencies or other parties in between who benefit from obfuscating or exaggerating reliability.

Infer ROI. DSPs, ATBs and SSPs use all of the directly usable data in a big-data approach where machine learning algorithms or employees try to discern patterns that increase ROI.

[65] KImon Zorbas & Ionel Naftanaila (28/01/2015, Brussels).

This means that one looks for patterns where inventory is bought for a low price and ensures a positive reaction towards the advertising campaign's goal.

Rocketfuel's Roland Siebelink provided an example to illustrate one of these patterns. Rocketfuel had to bid for Lexus, a luxury car manufacturer. Lexus wanted to increase leads for its dealerships in the US. A pattern emerged, Indian online newspapers sold cheap inventory to American visitors and these people responded positively to Lexus advertisements. Apparently Indian expats in the US visit their mother country's newspapers. Due to the time difference, the inventory is cheap for American visitors because it is Indian night time. The interest in Lexus is also tied to origin, Lexus is perceived as a status symbol in India. The inferred pattern is reverse engineered, Rocketfuel starts buying on multiple venues and after 50 to 70 successful acquisitions or other defined marketing goals, patterns begin to emerge. By buying more from successful venues and less from unsuccessful venues, the machine learns these patterns. The value of the data is only indirectly relevant, after these patterns have been learned.

Improve user experience. Personalising content, in most cases providing more relevant information.

Informatisation. As Mario Vanlommel[66] pointed out, much of their value for their end-users does not lie directly in their real time or historic traffic data, but an adaptation to the specific information needs of their customers: "make it high level and consumable or make it into the thing the organisation you are working for needs."

Reliability checks. Both Be-mobile and Twoo refer to reliability checks by comparing data from assumed similar occurrences. If something stands out, it is investigated before it is added or it is flagged as unreliable. Be-mobile, for instance, attaches a reliability score to incoming phone calls. Twoo prioritises more reliable complaints and this reliability factor is defined by the degree of information a complaining user adds.

[66] Mario Vanlommel (30/01/2015, Brussels).

5. Scenarios for the Instrumentalisation of the Shared Information Value

The research undertaken for this deliverable, and in particular the conducted expert interviews, discloses the ecosystem as it is today. The problems, which users face in the value network around personal data are central to the USEMP project: non-transparent re-use of personal data, often through unaccountable and untrustworthy third parties. This causes information asymmetries and power imbalances, disfavouring the user. Our research has shown that many actors on the industry side also see shortcomings in the status-quo, and are not per se contradicting the interests of users. The most desirable (and probably realistic) goal should, therefore, be to reconcile different interests through striving for mutual benefits.

A concrete and objective value of personal data is not possible to define. All actors depend to some degree on some kind of personal data, but each of them on different kinds and for different ends and purposes. Thus, its value differs for each actor. A mechanism as proposed by USEMP could increase transparency of data use in the ecosystem. Thereby, it would not only contribute to a business ecosystem that is more respectful to the content creators such as social network users. It would increase the value of (personal) data for each actor and thus for the ecosystem on the whole.

The low level of trust in the industry is a main barrier in reaching such a goal (van der Graaf and Vanobberghen, 2012). Trust is not only lacking from the user towards the industry (rightfully so in many regards), but also between all commercial actors in the value network. The reason for this is among others the non-transparency of data-related operations. Linked to non-transparency, the low quality of utilized data is a major issue, impeding the industry's functioning, reducing efficiency and thus also profits. For the user, non-transparency creates not only the information asymmetries, but also a decrease in user experience (for example, irrelevant advertising).

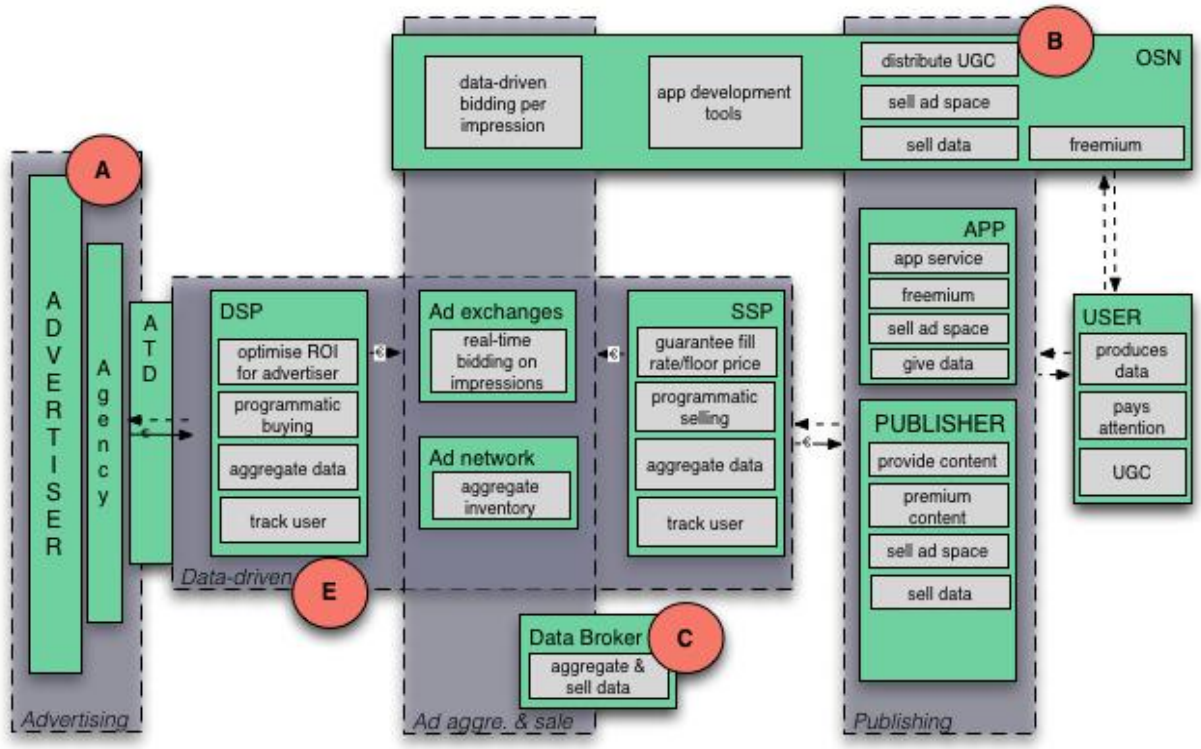

Figure 7: Where to locate the tool in the Value Network

5.1. As-is and As-would-be

The first iteration of the status quo needs to assess how the ecosystem will evolve, if the market is left alone, without appropriate regulation and mechanisms as aimed for by USEMP.

The low quality and reliability of data is a real problem for the industry. To avoid being responsible and accountable for personal data issues, actors often outsource the gathering and pseudonymisation of data to data brokers or similar companies. They purchase the user data they need for their operations from these third parties. This can be seen in a similar light as rights clearance in the field of intellectual property. After buying data in this fashion, the actor can reason that it is not anymore personal data, or if it is, at least lay the blame on the third party. This adds to the legal uncertainty, prevailing also due to the unclear provisions in the legislation.

As the interviews have illustrated, the functioning of the market is strongly affected by major OSNs, mostly Google and Facebook. These actors might be able to offer a remedy regarding data quality. They have full access to personal data through user profiles. In addition, they have access to pseudonymous data spanning over many data points: Facebook does that through their login tool and like buttons, Google due to the single user profile for its wide range of services. As a result, data is precise, qualitative and reliable. This elevates them to a central and very strong position in the value network.

Users might increasingly demand stronger data protection and prefer those actors and services they trust, that offer respective tools. If the market is left alone to deal with that, the already strong position of the prominent actors today will allow them to fortify their position. In contrast to smaller companies, these actors have the necessary means at their disposal to invest and develop new tools and mechanisms. So also data protection and privacy tools would be controlled by those companies, which already misuse their access to user data

today. This could be regarded in the light of feigned vendor relationship management, pretending that users have control.

If we evaluate the trajectory of developments in the context of personal data based on these aspects, strong companies with direct user ownership will be the winners. Such a development would not only hollow out the European market for data further and decrease its competitiveness. It would also render the personal data market even less regulable for European regulators. The following discusses how these unwanted outcomes might come about and how they might be circumvented.

The strive after data protection and privacy is understandable in the light of Google, Facebook and the business that is undertaken with user data in general. Still, our research indicates that maximal data protection and privacy through regulation can also lead to adverse effects; harming the European industry, weakening innovation, hurting SMEs and start-ups while favouring big companies, especially from non-European countries. The current data protection regime of the EU bears testimony to such fears.

The concept of privacy by design, incorporating privacy from the first step of the design project, aims at maximal privacy protection. Prior consent in the form of Opt-in is the prime example in this regard. User-centric data management in its purest form follows the same principles, enabling the user to take all decisions over his/her data.

The implementation of personal data management tools entails several challenges. Most significantly, the costs can be challenging for small companies. Efforts to ensure data protection, as demanded by legislation, are coupled with investments in respective technologies to support such tools. Cookie regulation sets an example in this regard, but also the right to be forgotten: the request to delete all data related to one user might be difficult to comply with for practical reasons rather than not being willing to do so.

Another decisive challenge regarding Opt-in is that it cannot work as those companies, whose operations are most data-driven, are located between advertisers and publishers. These companies do not offer any kind of value proposition that is relevant directly to the user. As a result, they have no leverage to interest users and to care sufficiently about their personal data. The most successful personal data companies of today, on the other hand, benefit several advantages. As social networks, they are able to provide opt-ins simple through existing user lock-ins. The value proposition is, at least partly, that the only alternative for the user to being tracked is to stop using the service. Their scale furthermore allows them to create independent data management tools, owned and controlled by them.

Thus, the value proposition of personal data management tools, i.e. the incentive to use them, is a central issue in this regard. First, it depends on the interests of the users not only in knowing what is happening with their data, but also in investing effort and time in actually controlling it. Second, even those groups in the society that are concerned about their personal data do often lack the knowledge to assess their own “value”. Expectation management would thus be integral to a tool in question, as an individual’s data is worth much less to the industry than many think. In addition, it needs to be clarified how much value a user already derives from using free services, which in turn are sponsored through the personal data market. These two aspects along could provide a good basis for a value proposition for a personal data management tool.

Third, such a tool could make clear that, if a user voluntarily provides certain data for instance through tracking, a company would make money of it, but also the user could

benefit of more personalized marketing. Due to pseudonymisation and implied Opt-in, there is no transparency of the choices of a user and the re-use of his/her data. This should be part of an adequate value proposition towards the user for using a personal data management tool. It might make people use it and, therefore, provide re-usable data for the benefit of several actors.

5.2. USEMP: A Centralised Tool

The final iteration of our findings assesses how we think how the USEMP tool for personal data management can become effective and interesting enough for users and businesses to utilize. The research conducted here has illustrated that direct monetization of personal data by the data subject is indeed rather unrealistic, due to the low value of an individual's information. Moreover, the value that users derive from free services, such as Google, Facebook, Twitter, news etc., is already immense, yet, indirect. Instead of monetization, a tool that focuses on transparency in the first place (with the potential to be extended) seems to promise most potential, both for the user and the business. USEMP's use cases and scenarios take into account the importance of transparency and awareness (see USEMP D2.1). Foremost, a tool should create knowledge and awareness about the personal data market, which in itself would already be a major step forward. Not only would it become more clear the actual value of personal data. It would also illustrate the implications of free online services. The development of visualisations, as provided for example in D6.3 of USEMP, is certainly of main importance in this regard, as it improves the user experience and comprehensibility of these complex dynamics.

From the value network perspective, which we have adopted for this work, it is a central question where such a tool needs to be "inserted" in the value network, in order to be most effective: which actor is most directly affected by the tool, and how can this actor pass the obligations and benefits to its partners in the value network. Due to the strong connectedness of the ecosystem in question, the right location in the value network can affect the whole structure. Depending on where the tool is located, different advantages may also arise for different actors, and for users.

The broad distinction we can make regarding the value network is between 'in the middle' and 'at the sides'. So far, regulators have prioritised actors in the middle (D in Figure 7 above). These are the most data-driven companies, and all user-related data might flow through their systems. However, they are only intermediaries, facilitators of the actions of others. Also these actors are obliged to implement consent mechanisms in their operations. This is probably the least effective spot to realize better data protection rules. Although the core of their business is data, and often pseudonymous data, they have no direct contact with the user. They depend also in this regard on the sides of the value network. Focusing on the actors in the middle is unfortunate, as they create value indirectly for the user, backing the free model by increasing efficiency and relevancy of advertising and other content. Without the data, they cannot work, as targeting, delivering and evaluating all depend on it.

The actors on the sides, i.e. publishers and advertisers (A and B respectively in Figure 7 above), are arguably a better location to implement a personal data management tool. The reason for this is that they have direct contact with the user or customer. They also need a good reputation to stay attractive for the user. Furthermore, almost all other actors depend on the sides. Thus, they do not only have leverage of the users, but also over the actors in the value network. Due to the strong connections and dependencies, business-to-business

pressure down the value chain should be utilized as a powerful accomplice in strengthening personal data rights. The user can only build a trust relationship with those actors he/she has direct contact with. These actors thus benefit of the trust from users, and because they interact and shape the rest of the value network, their strive for trust can expand.

As mentioned before, data brokers or third parties commoditizing play a role in the market of personal data, especially because other actors outsource the responsibility over personal data to these third parties (C in Figure 7 above). Therefore, one might argue to make them the actor that most directly interacts with the tool. However, they lack the leverage to interest users sufficiently. As they are dependent on the rest of the value network, they should also be affected indirectly via their business partners.

The most significant outcome of all possible scenarios thus seems to be that a centralized mechanism promises the most desirable outcomes; a tool with overarching effects on the whole ecosystem. This is not least the case because the value of personal data is indirect, and through centralising it, it becomes clearer. This can have positive aspects on both the user and the industry. Transparency is, after all, what is missing most in the ecosystem. Most importantly, a centralised tool could also promote smaller and European companies, which might otherwise not be able to invest in appropriate tools themselves or would be more affected than other, bigger firms. In addition to transparency, users explicitly listed trust as a requirement for such a tool, as the user research conducted for USEMP D4.1 has determined. Centralisation backed by official partners (e.g. the European Commission) would be a good starting point for creating trust on the side of users. Finally, a tool where all data-management related operations are centralised could create legal certainty and ensure accountability through transparency. The work conducted for the other deliverables of USEMP work package 3 will provide decisive insights concerning legal certainty of the tool.

In the course of our work, we have attempted to prioritise mutual benefits that might emerge from better and more user-centric personal data management. Centralisation promises several mutual benefits, and USEMP can provide such centralisation. The user will benefit of more transparency and of more control about data accesses, inferences and data-driven implementations. Here, the consent mechanism can also be improved, as a one-time consent (or disapproval) for a certain aspect could be implemented clearly all over the value network. In this regard, the work that is being conducted in course of USEMP work package 6 and presented particularly in D6.3 is interesting. The indications of personal data value to end users that are suggested there need to be taken into account for the USEMP tool, in addition to the aspects as determined in this work.

The industry would benefit, in addition to those points already mentioned, most importantly through increased efficiency. As was explained above, the ecosystem on its whole suffers under the lack of trust and the related low data quality. Applying the central tool and complying with the rules and standards it sets will then have positive aspects on the customer relationship, as well as on data veracity.

6.Conclusion

The personal data market is a complex structure of interdependent actors. The online advertising industry is an appropriate case study therefor. Moreover, it is advanced, influential and it is dependent on personal data. Despite the importance of the user, its role remains too passive in the value network. USEMP's user-centred personal data management tool intends to change this situation. Success for doing so depends on various factors, of which sustainability ranks far from lowest. The question of how to achieve sustainability is thus central to this deliverable.

An objective to this deliverable was thus to provide a sound starting point for establishing a tool that is sustainable as well as functional and innovative. The leading research question was designed in this light.[67] We have argued that it needs to be evaluated how (economic) value is created with data, directly as well as indirectly; what mutual benefits can be fostered with the USEMP tool; and where in the value network the tool can have most leverage. We can conclude first of all that the ecosystem has to be taken in to account in its whole as well as each of its parts.

We have commenced our analysis of these questions with a state of the art. It provides the value network of online advertising, the actors, their roles and relationships with a focus on (personal) data flows. The strong interdependencies in this connected, automation-based and data-driven ecosystem have become clear, as well as the disruptive nature of companies such as Google and Facebook. This research was complemented by linking it to the GDPR; the situation in the USA, where many personal data related operations originate; several existing tools and initiatives; and considerations about value propositions offered to users for providing their data, and for companies to use such data.

The empirical research, consisting of expert interviews, has added indispensable information. An important aspect here was also to assess incentives and costs for the industry to implement a USEMP tool. Main insights here were as following. Data (mainly personal data, but not only) is the decisive economic asset, because the industry is based on automation. For automation, data only works at scale. This means that pseudonomysed data is sufficient for the core operations, but also that there is no real value to the industry in the data of an individual. It can be argued that there is no objective value to personal data at all, as it is only valuable in relation to a certain purpose or the individual itself.

Possibly most important, however, the interviews have shown that the ecosystem suffers under certain shortcomings related to data, which are quite related to the user's drawbacks regarding the personal data market. There is no trust between the participating actors. This is the case mainly because also for them the value network and the other actors are not transparent. Directly related to the lack of transparency is low data-quality. These shortcomings directly effect efficiency for the whole value network. Therefore, also most of the industrial actors are hoping to revise them and improve their own operations.

[67] How can a personal data management tool be established that actually supports users in their providing of personal data, while being economically and strategically viable for businesses, i.e. sustainable? (see chapter 3.1)

Here, the potential benefits for users and businesses, as this deliverable set out to identify, seem to overlap most significantly.

The exploratory scenarios we developed in the final part of this work further illustrate that focussing on transparency, trust and increase in data quality are the aspects that seem to lead to the most strategically viable and thus sustainable path to be envisioned for the user-centred personal data management tool. This is also in line with other work packages of the project point in similar directions: the focus on advertising and transparency in the scenario on economic value awareness (see USEMP D2.1), the users requiring transparency and trustworthiness (among others) as established in USEMP D4.1.

A centralised solution for user-centred personal data management is the main recommendation that can be derived from the research at hand. The final scenario argues why this is the case; transparency, efficiency, trust and control. An initiative that USEMP should still pay attention to in this regard is Digitrust, as introduced above in chapter 2.5. It demonstrates that increasing efficiency for the industry is linked to increasing transparency for the user. Most importantly, the indeed worrisome position of actors such as Google and Facebook, not only in the ecosystem but in society in general, needs to be faced with appropriate solutions. Such should on the one hand protect the user, but on the other hand not discriminate against European and especially smaller companies. A centralised tool could provide a first step in the right direction in this regard and will be further explored in the second version of the initial and final exploitation reports (D9.5 and D9.7, respectively).

Bibliography

ANA. (2011*). Agency Trading Desks, Basics marketers need to know, questions to ask* (p.5).

Al-Debei, M. M., & Avison, D. (2010). Developing a unified framework of the business model concept. *European Journal of Information Systems*, *19*(3), 359–376.

Allee, V. (2009). Value-creating networks: organisational issues and challenges. *The Learning Organisation*, *16*(6), 427–442.

Article 29 Data protection working party. (2007). Opinion 4/2007 on the concept of personal data (No. WP 136) (p. 26).

Bandara, W., Indulska, M., Chong, S., & Sadiq, S. (2007). Major issues in business process management: an expert perspective.

Ballon, P. (2007). Business modelling revisited: the configuration of control and value. *Info*, *9*(5), 6–19. http://doi.org/10.1108/14636690710816417

Bouwman, H., Haaker, T., & Vos, H. (2005). Designing Business Models: A Practical and Holistic approach. Telematica Institute Enschede.

Cohen, N. (2008). The valorization of surveillance: Towards a political economy of Facebook. Democratic Communiqué, 22(1), 5–22.

Constine, J. (2014, November 30). Facebook, Google, and Twitter's war for app install ads. Retrieved January 27, 2015, from http://techcrunch.com/2014/11/30/like-advertising-a-needle-in-a-haystack/.

Chesbrough, H., & Rosenbloom, R. (2002). The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, *11*(3), 529–555. http://doi.org/10.1093/icc/11.3.529

Cookson, R. (2015, February 1). Google, Microsoft and Amazon pay to get around ad blocking tool. *Financial Times*. Retrieved from http://www.ft.com/intl/cms/s/0/80a8ce54-a61d-11e4-9bd3-00144feab7de.html#axzz3QnSOtgUU

Dubosson-Torbay, M., Osterwalder, A., & Pigneur, Y. (2001). Ebusiness model design, classification and measurements. *Thunderbird International Business Review*, *44*(1), 5–23.

Van Dijck, J. (2013). The culture of connectivity: a critical history of social media. Oxford; New York: Oxford University Press.

Facebook. (2013, April 10). Facebook Awards: Partner Categories, a New Self-Serve Targeting Feature. Retrieved February 11, 2015, from https://www.facebook-studio.com/news/item/partner-categories-a-new-self-serve-targeting-feature

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI Magazine, 17(3), 37.

Fowler, G. A. (2014, August 5). What You Can Do About Facebook Tracking. *Wall Street Journal*. Retrieved from http://www.wsj.com/articles/what-you-can-do-about-facebook-tracking-1407263246

Guest, G. (2012). Applied thematic analysis. Thousand Oaks, California. Sage Publishing

Hawkins, R. (2001). *The Business Model as a Research Problem in Electronic Commerce, Socio-economic Trends Assessment for the digital Revolution (STAR) IST Project* (Issue Report No. 4). Brighton.

Heyman, R. (in progress), The Commodification of Personal Information on Social Media. PhD, Vrije Universiteit Brussel, Brussels

Heyman, R., & Van Dijk, N. (2013). 3.3.1: Report on differences between user and legal perspective on privacy and profiling (p. 70). Brussels: iMinds-SMIT.

Introducing Facebook's Audience Network. (2014, April 30). Retrieved January 27, 2015, from https://www.facebook.com/business/news/audience-network.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. doi:10.1016/j.bushor.2009.09.003

Moore, J. F. (1993). Predators and prey: a new ecology of competition. *Harvard Business Review*, *71*(3), 75–86.

Ohlhausen, M. K. (2014). Privacy Challenges and Opportunities: The Role of the Federal Trade Commission. Journal of Public Policy & Marketing, 33(1), 4–9.

Osterwalder, A., Pigneur, Y., & Clark, T. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. Hoboken, NJ: Wiley.

Pfadenhauer, M. (2009). At eye level: The expert interview - a talk between expert and quasi-expert. In A. Bogner, B. Littig, & W. Menz (Eds.), Interviewing experts (pp. 81–97). Basingstoke [England]; New York: Palgrave Macmillan.

Schonfeld, E. (2008, September 14). LinkedIn To Launch Its Own Ad Network. Retrieved from http://social.techcrunch.com/2008/09/14/linkedin-to-launch-its-own-ad-network/

Siebelink, R. (2014, December). *The programmatic advertising revolution*. Talk given for iMinds-Smit-VUB in Brussels., Brussels.Al-Debei, M. M., & Avison, D. (2010). Developing a unified framework of the business model concept. *European Journal of Information Systems*, *19*(3), 359–376.

The Economist. (2014, September 13). Getting to know you. *The Economist*. Retrieved from http://www.economist.com/news/special-report/21615871-everything-people-do-online-avidly-followed-advertisers-and-third-party

van der Graaf, S. and Clippinger, J. (2012). Data Driven Urban Infrastructures and Trust Frameworks. White paper for the Institute for Data Driven Design, MIT Media Lab Hub.

Word Economic Forum. (2014, May). Rethinking Personal Data: Trust and Context in User-Centred Data Ecosystems. World Economic Forum Industry Agenda.

Welch, C., Marschan-Piekkari, R., Penttinen, H., & Tahvanainen, M. (2002). Corporate elites as informants in qualitative international business research. International Business Review, 11(5), 611–628.

ANNEX A) Interview Questions

Topic 1: Data

- What makes (non-)personal data valuable for the industry?
    - What are the different services and products derived from data?
- How much of (your company's/organisation's) revenue is created through data directly or indirectly?
- What are the different types of data used in your company?

Topic 2: Situate (your company/organisation)

Situate your company and the services it provides in the lumascape value network

- where does the data come from: collection/access (Also feedback data)
- how is data shared between entities?
- who stores it: who owns it
- who analyses it
- who uses it?

Identify the data streams and corresponding money-streams (concrete)

Name them

Value them: Costs & Benefits

Ask about technological means for collection, transfer, storage and manipulation

How is data treated?

- How is it anonymised?
- What is the reasoning behind gathering data?
    - As much as possible for big data analytics?
    - Data minimisation?

Topic 3: User centric data-management

- What is your experience with regard to user centric data-management?
    - How is this currently approached by (your company/organisation)?
- Is dynamic consent, the means for end users to control their data in real time, an interesting supplement to current data protection efforts?
    - Why?
    - Who should provide tools and manage dynamic consent?
- What practices currently enable data protection?
- What are the current challenges with regard to providing transparency?
    - Are more tools the solution?

Topic 4: The new GDPR

- What changes do you anticipate with the coming of the GDPR for your business?
    - Is your sector changing?
- Would this provide an 'incentive' to invest in transparency tools?
- What should be addressed in the GDPR?
- What should be addressed less?

End

A) What is interesting, but did not surface during the interview?

B) Can you refer us to other interesting parties to interview? may we refer to you as contact?

ANNEX B) Contacted experts and elite

Lien Brusselmans interviewed by Rob Heman, 21/12/2014, Gent. Lien is Marketing & Communication Manager at Engagor.

Roland Siebelink provided his insights during an event, 17/12/2014, at iMinds-Smit-VUB in Brussels. Roland is head of quality, productivity and best practice of Rocket Fuel.

Theodores Michalareas interviewed by Rob Heyman, 19/12/2014, via Skype. Theodoros works for Velti, provider of mobile marketing and advertising

Joelle Frijters interviewed by Jonas Breuer, 14/01/2015, Antwerp. Joelle is co-founder and CEO of ImproveDigital, a European provider of independent publisher monetisation technology. (improvedigital.com)

Chris Payne interviewed by Jonas Breuer, 21/01/2015, Brussels. Chris is Public Affairs Manager at World Federation of Advertisers. The federation is a global organization representing marketers and advertisers. (http://www.wfanet.org/en)

Niels Baarsma interviewed by Jonas Breuer, 23/01/2015, Brussels. Niels is co-founder and CTO of Yieldr, a demand side platform provider.

KImon Zorbas & Ionel Naftanaila interviewed by Rob Heyman, 28/01/2015, Brussels. Kimon is Director at the Digital Business Consultancy Group (www.dbcg.eu). Before he was CEO and Vice President for IAB Europe.

Mario van Lommel interviewed by Rob Heyman, 30/01/2015, Brussels. Mario is Technical Sales Engineer at Be-Mobile, a leading provider of traffic and mobility content and services for the automotive industry, mobile, media and government road operators.

Joost Roelandts interviewed by Rob Heyman, 06/02/1015, Brussels. Joost is COO at the social network Twoo.