Building future generation service-oriented information broker networks

A technical and legal joint perspective on the di.me case study

Sophie Wrobel
CAS Software
Karlsruhe, Germany

Mohamed Bourimi
MT AG
Ratingen, Germany

Eleni Kosta
TILT
Tilburg University
Tilburg, Netherlands

Rafael Giménez
BDCT Barcelona Digital
Barcelona, Spain

Simon Scerri
DERI NUIG
Galway, Ireland

***Abstract*—Future generation networks target collecting intelligence from multiple sources based on end-users' data and their social interaction in order to draw useful conclusions on enabling users to execute their rights to online privacy. These networks form a rising class of service-oriented broker platforms. Designers and providers of such network platforms during the design and development of their systems focus primarily on technical specifications and issues. However, given the importance and richness of user information collected, they should already at the design phase take into account legal and ethical requirements. Failure to do so, may result in privacy violations, which may, in turn, affect the success of the network due to increasing awareness with respect to users' privacy and security concerns, and may incur future costs. In this paper, we show how the di.me system balanced technical and legal requirementsthroughboth its design and implementation, while building a decentralized social networking platform. We report on our advances and experiences through a prototypical technology realizing such a platform, analyze the legal implications within the EU legal framework, and provide recommendations and conclusions for user-friendly service-oriented broker platforms.**

Keywords—di.me; online privacy; social media; software design; legal and ethical issues; broker platform; context-aware web services; user data

I. Introduction

Human beings in the modern, data-driven era are increasingly dependent on technology and systems to make information available for different purposes, with wide-ranging implications on society. Such technology needs to support transparent, conscious decision-making processes in order to earn (end-) users' trust and assist knowledge workers in gathering multiple perspectives and qualitative insights to form useful knowledge [1]. Popular online social networks (OSNs) such as Facebook and LinkedIn encounter difficulty in this area today in several respects and have been often criticized for establishing complicating user interfaces in order to discourage users from making informed choices about the handling of their personal information, which this paper addresses.

A. Challenges facing modern online social networks

To some extent the interests of software providers align with these of their users. The user wants to utilize software and the provider needs some amount of personal information to provide it. But besides the amount of data that is necessary and the restricted use of these data for a legitimate and obvious purpose, the commercial interest of providers contradicts the interests of individuals. From the provider's perspective, user information represents a valuable asset. Hence, providers and their commercial customers have a strong interest in collecting and processing more information about their users, e.g., in order to improve their protfolios, or to offer customer-oriented services. This counts especially for OSNs and web services in general, which have been a trend in the recent years. Many social networks and services are free of charge. Their business is at least co-financed by innovative exploitation and commercialization of the users' personal data [2]. As a result, design choices in OSNs reflect the provider's financially-driven goal of maximizing personal data exploitation.

In order to obtain and commercialize personal data, social network interface design has evolved to encourage data entry. Research has been done on everything from the optimal warning message color, to presentation layout, or auto-complete suggestions based on the information available about a user's friends [3]. As a result, the user is often encouraged to incrementally provide more personal information, often without fully understanding the consequences this may have on their digital identity due to a lack of digital literacy: even in the 14-49 age group, digital proficiency lies below 60% in most major European countries [4]. Concretely, digital literacy

The di.me project is supported in part by the European Commission under the Seventh Framework Program FP7/2007-2013.

involves skill and understanding of social networking, transliteracy, maintaining privacy, maintaining identity, creating content, organizing and sharing content, reusing and repurposing content, filtering and selecting content, and self-broadcasting [5].

B. The need to redesign for more privacy

OSN platforms and services play an increasingly important role in all private and business activities. Two of the key challenges facing OSN users with limited digital literacy are the implications of data transfer, and the rights they have on their personal information. These challenges extend beyond the realm of OSNs: they are equally applicable to the Internet of Things (IoT), or any other electronic data broker transmitting information between two online services or parties within a concisely defined context.

At the heart of these challenges lie data protection issues. European citizens have a right to protect their personal data, which can only be collected and processed for specified purposes and usually on a consensual basis. Moreover, they have the right to request information about all collected data about them, and the right to ask their rectification or deletion [6]. GéraldSantucci, Head of the Knowledge Sharing Unit at the European Commission’s DG CONNECT, writes, “How can we have the Internet of Things (or the ‘Internet of Everything’) while preserving our fundamental right to privacy? Several answers exist, but we have seen that they can actually be clustered around two: the first one is technology itself - embedding privacy and security in the very design of new systems and components; the second one is adequate rules and regulations. A combination of technology and regulation can also be a wise approach [7].” So, how can technology and regulation together effectively support such negotiation?

As Lessig argues, if law can regulate software, and software can regulate individual behavior, then software provides lawmakers with an effective way to shape the way their subjects behave [8]. Following that paradigm, the software provider has a responsibility to ensure that data protection requirements and other privacy obligations imposed by legislative institutions are technically supportedbyconsidering them early in the design of the respective systems. The service provider has a responsibility to ensure that collected data is handled securely and appropriately, and network providers have a responsibility to ensure that communication channels are protected, while users are responsible for their conduct. Such responsibilities should be taken into account already at the design phase of systems and applications, in respect of the ‘privacy by design’ principle, which enlists technology to protect individual privacies by default on a continual basis [9]. By introducing law and ethics as core values during the software requirements and design phase, the resulting implementation could provide a solution that does not conflict with the right to privacy, or with exploitation interests (to some extent for current requirements)[1].

This paper considers how introducing these core values during the requirements and design phase of di.me resulted in a privacy-oriented service-oriented architecture (SOA), which has the potential to intelligently assist, without restricting, a safe and deliberate participation of less digitally literate individuals in popular OSNs. It describes the di.me context and architecture, and analyzes how di.me reacts in select use cases against critical data handling concerns that are commonly expressed against popular OSNs.

This paper focuses on the legal requirements relating to data protection affecting software design. The service provider and network provider layers lie out of the focus of this paper, although they are also affected by data protection regulation [10].

The remainder of this article is structured as follows: while the current section motivated the problem statement, the next section addresses di.me as a case study in this respect. Section III compares our contribution to related work. Finally, Section IV concludes our contribution and outlines potential future directions.

II. The Di.Me Case Study

di.me is a distributed OSN, which additionally serves as a personal information broker platform. It operates as a digital identity management tool, allowing users to maintain an overview of their data across various supported online services, such as LinkedIn or Twitter.di.me operates as a privacy-enhancing technology (PET) platform, by intelligently warning users when their online interactions involving data may lead to undesirable consequences. It also operates as a data exchange broker, by allowing users to share personal data with other online services in a secure and safe manner. By considering legal and ethical values during the requirements phase of di.me, as well as the subsequent system and component design and implementation, the result is a privacy-oriented information broker platform, which negotiates between di.me users and other OSNs to enable free-choice and context-specific data transactions [11].

A. Situational description of di.me

To describe how di.me operates and the issues it solves, consider a series of illustrative scenarios revolving around a typical modern individual, Alice. These scenarios will be treated from both a technical and legal perspective in the following discussion.

1) Multiple digital identities

Alice acts differently under different situations. For simplicity, consider two roles which Alice fulfils: (1) *business*: on a business trip, she meets a new potential customer, Bob. They exchange business contact information, and Alice invites Bob for a dinner conversation. During dinner, they make a verbal commitment on a business partnership. The following day, Alice sends Bob a sales contract. (2) *friend*: taking advantage of the travel opportunity, Alice does some sightseeing. She meets a friendly lady, Carol, at the beach, and excitedly posts about it on Twitter. They befriend each other on Facebook, where Alice posts pictures of their beach trip, and promise to stay in touch.

[1] Future or change requirements, e.g., after lab and user trials could result in changes that need redesign and retrofitting of the current implementation. The following Sections describe therefore our contributions for the current design and implementation of the di.me userware.

In di.me, these multiple digital identities are embodied in the form of profiles. A profile is a set of information about a user that she provides to other users or services. A di.me user then gives other persons and groups access to her profile by sharing a profile card with them. While a profile card does not contain any information itself, it is a context-specific access token allowing a particular recipient to retrieve the associated profile information [11].

2) Intelligent context recognition

Personal devices can be used to determine where a user is and what she does. Suppose Alice carries her mobile phone with her constantly, and does most of her work on her company laptop. When she is connected to di.me from her company laptop, there is a very good chance that she is working. Personal devices are just one contributing data type in Alice's context (e.g. geo-locational, attentional, nearby peers, environment conditions, IP address, etc.), from which her situation can be deduced – for example, whether she is actually working, or whether she is hanging out with her friends after work.

In di.me, the user's context can be deduced by considering the live contextual information stemming from her devices (e.g. desktop, mobile device). Each device has always one dynamic live context. Snapshots of this live context are saved as static situations [11].

3) Information flow management

Alice is an active digital community member: she shops online, pays through online banking, and even has digital health care records and smart utility metering. But these conveniences aren't always as convenient as she would like: Every time she visits a new online shop, she needs to fill in all registration information all over again. And with each online shop or online social network having its own terms and conditions in what the end user often perceives ascryptic legal language, she can't be bothered to read through them every time. On the other hand, she often finds herself wishing that some registrations could take place automatically – for example, automatic registration at all baby bonus programs at her favorite stores when her child's birth shows up on her digital health record. However, after a few purchases, she starts to receive invitations and advertisements on baby products from companies that she has never heard of, and has no idea who might have given them her contact information.

di.me wishes to tackle many problems from the privacy and legal common point of view. One of these problems is concernedwith data transfers without the knowledge of the users. It acts as an information broker by allowing Alice to share her information with the parties she wants to share it with, while warning her if she inadvertently tries to share her information with parties that she may not want to share her information with [11].

4) Broker platforms in the digital landscape

Today, there are many platform solutions specializing in the sphere of contextualized information. Some focus on connecting information from entities, characterized as "big data", and others focus on connecting people, often called "social media". But these two trends are closely connected to tosome extent: they both deal with sharing contextualized information, which gives rise to service-oriented digital intelligence – a space in which broker platforms assist users to achieve meaningful information connectivity that is not addressed by popular market solutions: a mediating platform that can connect between data-driven platforms with people. As illustrated in Fig. 1 [12], this is a field that is largely unexplored by mainstream commercial offerings, but also a field which will flourish as a natural next step in internet connectivity.

Broker platforms in a SOA approach, like di.me, allow people-centric platforms to communicate with technology-centric platforms [13], while restricting data processing for a particular purpose in a defined context.Data brokerage in a service-oriented internet needs to consider not just technical but also legal implications, and define and negotiate responsibilities appropriately across multiple involved parties, including the user, the service provider, the network provider, and the software provider. While di.me itself does not facilitate negotiation, it does facilitate controlled data transfer in a user-centric way.

B. Technical description of di.me

The implementation of the di.me platform prototype technologically enables personal data usage in a controlled, trustworthy, and intelligent way [14]. It specifies a platform incorporating user-control deeply in design: a personal server (PS) that enables a di.me node in a decentralized network to connect to other users' PSs or external services, like various social networking platforms as mentioned above, and this by using distinct identities [15]. This node integrates all personal data in a personal information sphere, including user interests, contact information, files or resources, and social network services. Intelligent features and PETs further guide user interactions with the digital sphere, illustrated by context-aware access control [16], trust and privacy advice, or organizing their personal information sphere [14]. Besides integrating existing networks and services, the platform provides its own OSN functionalities, which are not available in known and popular OSN, in particular network anonymity [17][18][19].

1) Semantic model: information classification in di.me

The di.me Ontology Framework,based on the Personal Information Model (PIM) Ontology[2], is a differentiating concept allowing di.me to react to users with multiple digital identities, multiple use contexts, and differing objectives when sharing information. Each person in the di.me network owns a PS and an associated Research Definition Framework (RDF) store that contains the PIM representation. Amongst other information, the PIM includes references to persons, groups, service accounts (DAO), devices (DDO), resources (NIE), profiles (NCO) and live posts (DLPO). The PIM is extended byprivacy preferences (PPO instances), which enables the representation of databoxes, profiles and whitelists/blacklists, privacy and trust levels(NAO), andcontext information (DCON instances), which includethe unique live context representations of situations [20][21].

[2] Ontology descriptions are available underhttp://www.semanticdesktop.org/ontologies/ with exception of the PPO, which can be found under http://vocab.deri.ie/ppo

← Information Connectivity →

Data-driven Intranet *Connecting Knowledge* • Internet of Things • Semantic Search • Open Data Alliances	**Service-oriented Intranet** *Connecting Intelligence* **<Future: di.me and other broker platforms>**
Static Intranet *Collecting Information* • Traditional Websites	**People-centric Intranet** *Connecting People* • Social Networks • Communities • Talent Markets

← Social Connectivity →

Fig. 1. Four quadrants of internet platforms for technology-mitigated information connectivity [11].

This extended ontology set, depicted in Fig. 2, combines information from personal and contextual spheres, which together with the trust and recommendation engines allow di.me to identify context recognition as well as to derive privacy recommendations.

2) System context: deriving contexts but protecting identity

di.me's global architecture follows a decentralized approach emphasizing near real-time asynchronous network interoperability, data-centrality, and user control. The PS, hosted in the Personal Server Layer (see Fig. 3), is the central element in the system architecture, being responsible for collecting, safeguarding and managing the entire user's data.

Client applications triggered from user's personal devices provide light-weight user interfaces to access the PS. Communications between the personal devices and the PS pass through a proxy layer to minimize traceability. The personal server is responsible for holding the user's information and providing computational capabilities, and can be securely deployed on the user's personal devices, on trusted commercial hosting services or in a hosting service provided by the di.me system. These concepts are well-aligned with those being pushed today by relevant initiatives within the distributed social networks scenario [22].

The wide range of devices allow for di.me to derive contextual information: Usage of a certain device in connection with particular users or a particular location can imply a particular context. For example, Alice sharing a document from her laptop connected from her office IP address implies that she is probably in her 'business' profile in an 'at work' situation. In order to protect Alice's identities from being traced back to her, her requests are routed through the di.me proxy layer.

3) System architecture: powering smart recommendations

The PS itself comprises of multiple components which work together to provide intelligent analysis of identity and context information provided by the clients. Its high-level PS internal architecture, shown in Fig. 4, is related to that of dynamic webapplications [14], and also favored by the separation of the addressed concerns inherent to the multi-layersystems.

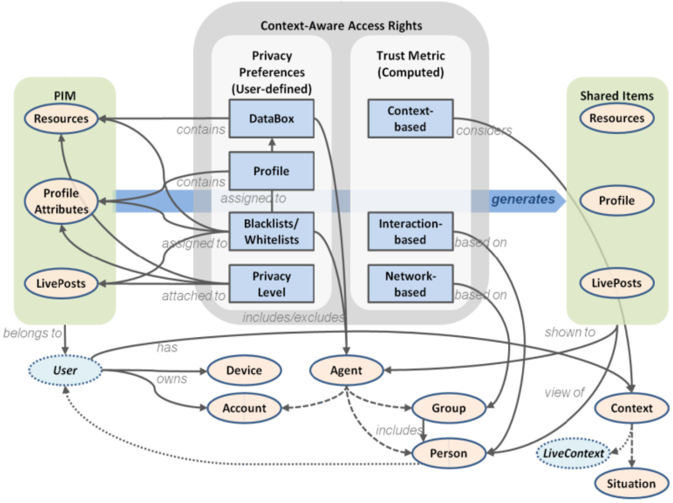

Fig. 2. High-level di.me semantic model description.

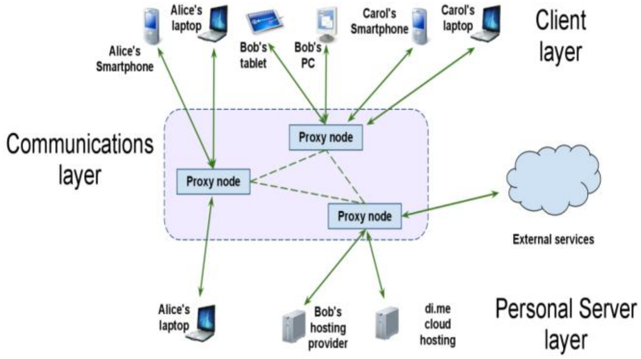

Fig. 3. Global di.me system architecture schema in a large-scale deployment.

Within this approach, the persistence layer isolation also benefits the decoupling from the underlying database technology and enables a multi-engine deployment. This feature is especially useful for the di.me system, intended to store heterogeneous data with fairly different access requirements such as the user's personal data, context data or service crawling schedule information.

The semantic and storage modules are used to store semantic as well as environment data. Semantic data includes information required for semantic deduction, such as 'the beach' – which could be a potential location nearby. Environment data includes information required for system operation, but without a semantic value, such as the strengths of the nearby wifi access points. In addition, the semantic module crawls connected web services to retrieve associated data at a pre-defined refresh interval. For example, Alice can connect to twitter, and the crawler would retrieve tweets, profiles, and friends and followers once an hour. The contextprocessor module derives contextual information from environment data.

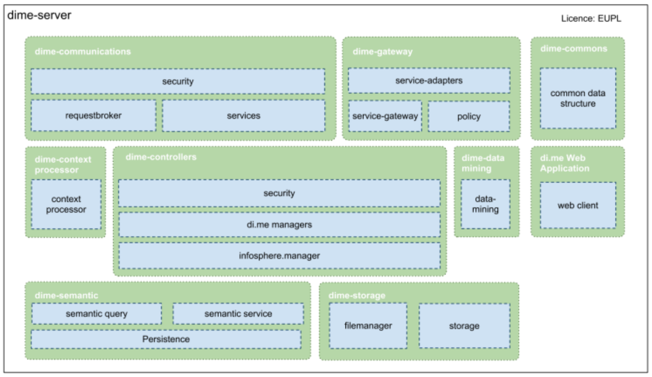

Fig. 4. Multi-layered architectural model for the personal server.

For example, based on how strong nearby wifi signals are and what the nearby wifi network access pointnames are, the contextprocessor can determine when Alice is in the vincity of her office network. The datamining module derives an adaptive privacy score to persons and to data indicating how trustworthy that particular resource is and checks whether data updates trigger a warning. The privacy score is calculated according to the di.me trust model which accepts inputs from the di.me semantic store and outputs a probability score, which adapts over time with respect to the user's interaction patterns. The gateway module manages and transforms communication entering or leaving the personal server with relevant policy rules. For example, a 'no twitter at work' policy would prevent her from posting to twitter if she was in an 'at work' situation [21].

When Alice posts "Sitting on the beach with @carol!" on Twitter through di.me, di.me's semantic analysis of the message recognizes an activity (sitting), a location check-in (beach), and a person (Carol). Triggers set by the controller module in the datamining module are fired, as the combination of personal information relating to third parties (Carol's identity) and their common location (beach) is a potential privacy issue, and this causes di.me to present Alice with a warning message informing her about that risk, and asking if she is sure she wants to post [14]. Unlike popular applications where users are expected to have these digital literacy skills, di.me allows non-literate users to participate in online social networks while informing them of risks only during relevant situations and thus minimizing the likelihood that the warnings get ignored. The final layer is the authentication and authorization layer, which ensures that all transactions are only honoured when the credentials are valid.

C. Legal perspectives

In order to ensure the protection of individuals, the European legislation on data protection applies when the processing of personal data takes place. The data can be processed only under the grounds mentioned in the Data Protection Directive [22] and their processing has to respect the basic data protection principles. The obligations stemming from the data protection legislation have to be taken into account already from the designing phase of systems and applications ("privacy and security by design") [23].The European Data Protection Directive is currently under review. In January 2012, the European Commission presented its proposals for the reform of the data protection legal framework of the European Union, proposing the replacement of the Data Protection Directive with a Regulation, which was the outcome of consultation and debates of three intense years [24]. The proposed Regulation dedicates an article to the principles of data protection by design and by default[3]. According to this principle, both at the time of the determination of the means for processing and at the time of the processing itself, a controller must implement appropriate technical and organisational measures and procedures in such a way that the processing will meet the requirements of the Regulation and ensure the protection of the rights of data subjects.

Before moving into the detailed analysis of di.me from the legal perspective, a short introduction must be made to the terminology that is relevant for data processing operations. The term 'personal data'[4] is defined as 'any information relating to an identified or identifiable natural person ('data subject')'; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental economic, cultural or social identity. As regards the phrase 'identified or identifiable person', the possibility of matching data processed by a computer to a specific person will depend on a number of factors, such as who is doing the matching and what their technical capabilities are, what type of data is involved, whether other data are available to aid the matching etc.

'Data processing' is defined as "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organisation, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction"[5]. It follows that the definition of processing is very broad, so that it is difficult to conceive any operation performed on personal data, which would not be covered by it. It is important to note that even the mere storage of personal data constitutes 'data processing', so that simply storing data on a server or other medium is deemed to be processing, even if nothing else is being done with the data.

The relative data protection legislation defines three distinctive categories of parties:

- *'Data subject':* the individual to whom personal data refer to.
- *'Data controller':*an entity which alone or jointly with others "determines the purposes and means of the processing of personal data"[6]
- *'Data processor':* a third party who simply processes personal data on behalf of the data controller without controlling the contents or use of the data.[7]

[3] Article 23 of the draft Regulation

[4] Art. 2(a) Data protection directive

[5] Article 2 of directive 95/46/EC [18], hereafter called Data Protection Directive.

[6] Article 2 (d) Data Protection Directive

The classification of an entity as ‘data controller’ or ‘data processor’ is of great importance, for several issues, such as who shall carry the obligations appointed to the ‘data controller’ by the Data Protection Directive and who is to define the details of the data processing. As a rule of thumb it can be said that the data controller is liable for violations of the Data Protection legislation, while the role of the data processor is reduced.

Under the regime established by the Data Protection Directive, a key concept is that of ‘data subject’s consent’. If the data controller obtains the data subject’s consent then he/she is broadly free to process the personal data. The Directive defines ‘data subjects’ consent’ as being freely given, specific and informed[8]. It supplements this in the substantive provisions when referring to consent as being ‘unambiguously’ given[9]. Indeed, the definition of ‘consent’ in the Data Protection Directive is quite restrictive, requiring that the data subject be clearly informed in advance of what he is consenting to and that any processing of the data going beyond what is disclosed to him will be deemed not to have been consented to, meaning that it will be invalid. Particular risks arise in the online environment since there is an increased danger that the data subject might not have been fully informed or might not understand exactly what he is consenting to.

The related EU FP6 funded PRIME project relates to a privacy and identity management system that was demonstrated through collaborative E-Learning and Location-Based Services (LBSs). This differs from a broker platform in that its scope is more heavily directed towards inter-service connectivity, and LBSs are just a subset of potential di.me contextual entities. PRIME developed a set of requirements for Identity Management Systems (IdMSs) translating the obligations of the data protection legislation into requirements for IdMSs [25][26]. The PRIME requirements list has been used for di.me, which considered them in its design and development process.

To illustratehow di.me addresses issues surrounding the ethics of data transmission and user privacy today, consider several relevant di.me API and behavior around some critical ethical concerns around data handling within the scope of the previously described scenarios:

1) Linkability: Transfer of data to other contacts

di.me respects and safeguards user privacy by using strong, secure pseudomization techniques[15][16][17][18][19]. Because di.me acts as the intermediary and not the end service[10], it is not possible to make the legal analysis very concrete on this aspect. However, it is an important aspect of ensuring that users can exercise their right to digital privacies.

Pseudonymity: di.me uses the *idemix*[11] library to create secure credentials for information exchange between profiles. *idemix* allows the desired pseudonymous credential exchange, while still offering the possibility to de-anonymize user pseudonyms when needed – such as in the case of abuse or for accounting purposes, as required by law enforcement or for financial transactions [27]. This enables di.me to operate by transmitting personal data only via secure credentials, and on a completely pseudonymous basis, which is critical in ensuring that multiple identities managed from one central point can be unlinkable in all information flows within the di.me environment and this at least at the technological level [17][18][19].

Data exchange profiles: Each user can adopt and manage multiple public and private digital identities [15], which can be completely unlinkable if he strongly adopts *idemix* as an anonymous credential system at the level of personal attributes, with special attention to shared attributes across different identities. This separation of profiles allows a clean separation of business and private data, and which private data is shared with which business.

Trust metric [14][16][28][29]: An adaptive user trust index allows warning messages to be displayed, preventing a user from unknowingly sending a confidential file to the wrong audience. This concept is also applicablebeyond the di.me prototype no interactions in and between social circles, such as friends or business contacts. The di.me trust metric is calculated based on several inputs, illustrated in Fig. 5, including:

- *pre-defined trust dimensions:* When Alice uploads a photo in di.me to share with Carol, the photo is automatically given a privacy value of high. She can change this if appropriate.
- *recognition of user context:* When Alice shares her photo, di.me recognizes that this is a potential risk situation.
- *previous interaction:* The trust model uses available information from the semantic engine about the sort of information Alice has shared with Carol in the past, the situation, purpose, and context under which Alice is in now, and the current privacy value of both the photo and of Carol in order to calculate a probability value for the risk involved.

2) When a risk is identified, this generates an advisory, which is presented in the user interface, and Alice sees an advisory asking whether she is aware of the privacy risk involved in sharing her photo.Tracking context in information sharing

When you share information in di.me, di.me reveals personal information relevant to the share (See Table I).Note that each information share is associated with a *saidSender*. This *ServiceAccount* is a representation of a unique combination of a particular profile card and a particular web service account. The profile card is an access ticket to a set of information, available at downloadUrl upon presentation of appropriate access credentials, with respect to a particular context.

[7] Article 2 (e) Data Protection Directive

[8] Article 2 (h) Data Protection Directive

[9] Article 7 (1) and 26 (1) (a) Data Protection Directive

[10] End seviceper definition from legal point of view

[11] https://prime.inf.tu-dresden.de/idemix/

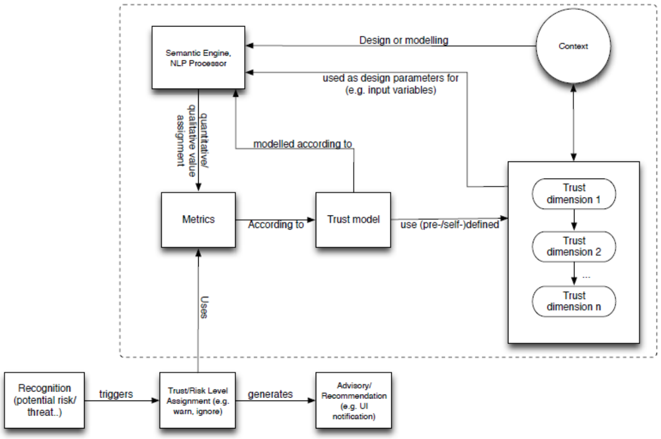
Semantic Engine, NLP Processor
Design or modelling
Context
used as design parameters for (e.g. input variables)
quantitative/ qualitative value assignment
modelled according to
Metrics
According to
Trust model
use (pre-/self-)defined
Trust dimension 1
Trust dimension 2
...
Trust dimension n
Uses
Recognition (potential risk/ threat..)
triggers
Trust/Risk Level Assignment (e.g. warn, ignore)
generates
Advisory/ Recommendation (e.g. UI notification)

Fig. 5. The system context of the di.me trust model.

When Alice shares her sales contract to Bob, this sharing is done through her business profile card. This contextual information is stored in di.me and is visible together with all other information related to the file or person involved in the sharing. Further, based on this interaction, di.me associates both the sales contract file and Bob with Alice's ‘business’ role. If Alice tries to share the sales contract with Carol, who is associated with her ‘friend’ role, di.me warns Alice: Carol is associated with Alice's ‘friend’ context, but not her ‘business’ context, and this would give Carol insight toAlice’s ‘business’ profile, which she does not yet have. Alice can then decide if the action was inadvertent, or whether the action was intentional and approve it. This functionality of di.me aims at protecting the privacy of users and raise awareness with regard to the sharing of their personal information. Although in principle users have the right to share their personal information whenever and with whichever entity they wish to, very often they do not realize that they are actually sharing personal information. di.me does not create profiles based on users personal data and context for any other purpose but to enable users to control the sharing of their personal information in an easy and comprehensive way. In this way it enables users to control better the information about them that they are sharing.

3) Designating purpose in information sharing

The second component of the ServiceAccount used for sharing information is a web service account. This is different for sharing between di.me users and sharing with other social networks. For example, when Alice shares the sales contract file to Bob directly using di.me, di.me creates a unique adapter for each contact’s profile that she shares to. More concretely, if she knew Bob as both a business partner and a friend, she could share the sales contract to Bob the business partner, to Bob the friend, or to both: di.me creates one web service account for each of the relationships she has with Bob, and this allows di.me to build an overview of the purpose associated to the data sharing by tying the purpose to the profile-specific web service recipient for information sharing.Alice is asked to choose which profile of Bob she wishes to send the file to and in this way she is offered the possibility to keep personal and professional information separate. This functionality actually assists Alice in determining the purposes for which she wishes to use specific information.

TABLE I. API DEFINITION FOR SHARING

Type	**POST /api/dime/rest/<user>/resource/@me**	
	Field name	***Description***
timestamp	created	When the user shared the information
string	downloadUrl	URL to access the shared information
GUID	Guid	List of service-specific configuration settings
string	imageUrl	URL to obtain a thumbnail of the shared information
timestamp	lastModified	When the information was last modified
string	Name	Name of the resource being shared
GUID	saidSender	Service account ID of the sender
List <GUID>	groups	List of groups to share with
List <GUID>	persons	List of persons to share with
integer	privacyLevel	Privacy level of the files being shared
integer	fileSize	Size of the file to be shared
string	mimeType	MIME type of the content being shared
string	Type	What is being shared. Valid values: resource
URI	userId	The user who is sharing (@me for current user)

On the other hand, when Alice shares her beach pictures to Carol by posting them to Facebook, this sharing is done under a generic information processing purpose as defined in Facebook’s user agreements. While this does not allow a specific data purposing to be shared over Facebook, it allows technical support for popular platforms which do not allow for specific data purposing in their API. To mitigate the damage that this could do, di.me’s architecture includes a multi-dimensional policy matrix which allows service providers, corporations or users to enforce desired technical guidelines, such as ensuring that collected data is consistent with minimum data collection requirements, providing a default trust metric value for information for a particular data source, or ensuring that predefined combinations of outgoing data are blocked. These functionalities of di.me enable service providers and corporations to comply with the data protection legislation. Allowing for the collection of only adequate, relevant and non excessive data in relation to the purposes for which the data are collected or further processed is a fundamental data protection principle, commonly known as ‘data minimisation principle’. By warning Alice when the data she is sharing is not consistent with the context she is sharing in, di.me assists Alice in protecting her data.

4) Erasure of data

Very few popular OSNs support data deletion, although they just may support hiding old data from the user’s visible experience[12]. As such, information shared via external services

[12] Providers must retain data for a specified duration as specified by data retention laws.

may not allow data deletion, and di.me cannot change this. For sharing inside of the di.me network, however, there is a mechanism to revoke access to data: the data is available only via the shared *downloadUrl*. If the di.me personal server hosting the *downloadUrl* happens to be offline, then the data displayed is shown from cached values that are updated at the next successful regular crawler synchronization, at which point the old values are updated with the current information available at the *downloadUrl* – which could include that the data has been deleted. The result is that when Alice removes the ‘phone’ attribute from her ‘friend’ profile, Carol will not be able to see Alice’s phone number anymore. di.me enables users to erase their data, without requiring any activity from the party that holds their information. In this way di.me provides an advanced functionality allowing the users to exercise their right to erasure of their data.

5) Control over data

di.me crawls data from all connected services on a regular basis and stores this data in its semantic store in order to provide the user with context-specific trust and privacy warnings. This data is crawled at regular intervals and refreshed, replacing old data from connected services. It does not broker data to third parties without explicit consent; each user can only share his own personal data with other services. However, di.me can be operated in single-user and multi-user modes. In the single-user mode, a single user runs the server and controls the data on the server for private use. The more controversial scenario is the multi-user mode, in which multiple users share a single di.me server instance. Each user still only has access to his own data, but the data is stored on one communal infrastructure. di.me allows users to have full control over their data when it is operated in single-user mode. When di.me is operated in multi-user mode, profiles are still maintained separately: there are no common profiles even if two data owners share a mutual contact. This allows di.me to ensure data set access in the same manner as when operating in a single-user mode. In this way, di.me enhances the transparency of the transactions and allows user to remain aware of any data sharing that relates to them.

6) Data monitoring

di.me crawls connected external services on a regular basis and alerts the user of substantial changes. For example, if it detects that Alice has befriended Bob on Facebook, and that there are so many similarities in Bob’s data on Facebook and her di.me contact Bob, di.me makes a recommendation that you merge Bob’s profiles to be associated as the same person. This construct would mean that Alice knows Bob in two contexts: as a ‘friend’, but also in her ‘business’ profile. These recommendations allow Alice to structure her contacts in a more organised way, and facilitate the sharing of her information in a more efficient way depending on which of Bob's profiles she wishes to send the information to, as described above.

7) Exercise data subject access right

di.me provides an overview of services that users are connected to. Each service is described in di.me as a *ServiceAdapter,* which is described in Tables II and III. The most important descriptors here are the *SAdapter.Description* field – which provides the user with a description of what the service is intended to do, and what connecting the service willbring as a benefit – and the *SAdapterSetting* definitions. One critical instance could be a Privacy Statement document included as a mandatory link, to which the user must agree, and is enforced when a service connection is built. When the service adapter connects to a service which displays the terms and conditions during the authorization protocol (as OAuthservices do), di.me does not need to include this as a mandatory link, but for services without such protocols (as services using basic HTTP authentication), this inclusion is critical. The privacy statement document ensures that the user consents to the collection and processing of the clearly defined purposes of di.me, namely to:

- Exchange and share profiles, messages, and data
- Provide the user with full control over who gets access to which information
- Allow the user acces via internet or the Android applicaton‘di.me mobile’
- Manage data from different user devices
- Enable to connect to information from other social networks (e.g. messages, liveposts, profiles, or contacts) and to update this information regularly
- Provide recommendations on data privacy and trust
- Analyse the situation of the user (e.g. to show which contacts are located nearby)

TABLE II. API ANNOTATION FOR SADAPTER (SERVICE ADAPTER)

Type	**SAdapter**	
	Field name	***Description***
URI	authUrl	URL at which credential exchange takes place
boolean	isConfigurable	Whether the service can be configured or not
List<SAdapterSetting>	settings	List of service-specific configuration settings
string	description	Description of the service
URI	userId	User associated with the service

TABLE III. API ANNOTATION FOR SADAPTERSETTING

Type	**SAdapterSetting**	
	Field name	***Description***
string	name	Description of what the setting is
enum	fieldtype	Possible values: boolean, string, password, account, link
boolean	mandatory	Whether the setting is required or not
<mixed>	Value	User-provided setting value

A similar mandatory Data Subject Access Link could provide information about where data requests can be sent. This is included by default for each service in the suggested di.me configuration files and labeled ‘You can request your data from <link>’. This link allows the user to exercise his right to request and retrieve information about his personal data, when they have been transferred through the di.me system. In this way di.me facilitates the exercise of a cornerstone data protection right of the users. All of these important links are then displayed in an easily accessible form in di.me in the service overview screen.

di.me's own data can be exported through calling the /api/dime/rest/<user>/dump API call, which provides a copy all the data that di.me stores on <user>. This authenticated call is only accessible for the user himself.

III. Related Work

Due to the multi-disciplinary nature of this contribution, note that this article is a summary of three years of research and design activities within the di.me consortium, which is constituted by nine partners from different countries across Europe. The project considered requirements categories in order to balance research and development outcomes in a multilateral manner [11]. The cited literature in previous sections reflects these outcomes throughout the project duration: trust, privacy and security were considered throughout the project, and in this order[13] for trust metrics and advisories [14][16][28][29], anonymity and secure communication [17][18][19], while considering unlinkability in the case of multiple identity support in a decentralized OSN [15]. The focus of this article, however, remains on how these numerous contributions are aligned with legal and ethical issues.

Building on results of projects such as PRIME[26], PrimeLife [14] and PICOS [15], incorporating leading privacy-oriented design methodology models such as privacy-by-design [9], and considering ethical perspectives expressed by contemporary media theorists [5][8][13],di.me demonstrates that a strategic privacy-oriented approach to social networking is feasible. di.metakes the data protection principles that are included in the European Data Protection Directive into account, and ensures the rights of the users.However, the pure technical consideration of technologies such as PETs is not enough to assess the consideration of all requirements from the legal and ethical points of view. There are many trade-offs (e.g. between privacy and context awareness) that could result in violations. For instance, since di.me supports multipleidentities, it was crucial to integrate unlinkability support in it. From a software engineering perspective, linkability as non-functional requirements (NFRs) may conflict with other competing NFRs such as providing context and collaboration awareness[16] at the user interface level, or negatively affecting user experience in terms of performance penalties by using anonymity networks.

Furthermore, there were many parties involved within the consortium and all requirements had to be considered from the legal point of view. For this, requirements negotation, elicitation, alignement, and priorization support necessarily occurred at process level. In order to address such complex cross-functional integration issues [30], the AFFINE methodology[17] [31] was followed within some workpackages in order to facilitate multi-lateral requirements cross-functional integration.Indeed, a complex analysis of all requirements by involving different partners with different goals and assessing thereby the correctness of design and implementation of agreed requirements can not be just solved by using various PETs (as demonstrated in [15] and solve in [18] and [19]). For instance, AFFINE enforces the earlier consideration of multilateral security requirements along with other (N)FRs also by involving all stakeholders, negotiating and aligning their potentially conflicting interests in the design [18] and development process, which meets our argumentation for privacy-by-design according to [7] and [8] in previous sections.[19]

IV. Conclusion

Introducing law and ethics as core values during the requirements and design phase of di.me resulted in a distributed OSN implementation that does not conflict with the EU right to privacy, and is also not contrary to exploitation interests of potential network operators. The resulting di.me prototype demonstrates that technology and regulation can work together effectively to support data access negotiation, and offers a protection mechanism for the less digitally-literate by presenting them with warning messages only in relevant scenarios, which make conscious and informed decisions concerning the potential repercussions of their interactions in and around OSNs. Although the prototype itself does not have the critical mass of users to become a replacement for current popular OSNs, it presents a concept that those OSNs could adopt, should they be required to.

Policy makers shape technology, and technology, in particular software, shapes user behavior. With American technology companies operating the vast majority of popular OSNs, the way European users of OSNs behave is slowly being shaped by this technological choice. But European policy makers can shape technology, and so requiring technology to

[13] The reader may excuse the emerging impression that the authors are citing their own work more than necessary. For accuracy, we cite these contributions since they represent sub-contributions in the involved research areas of security and privacy, data mining and linked data, usability engineering, etc.

[14] http://primelife.ercim.eu/

[15] http://www.picos-project.eu/

[16] Social, group, and workspace awareness answering 'who' is collaborating with 'whom', 'where', 'when', and 'why'.

[17] Agile Framework For Integrating Nonfunctional requirements Engineering is a Scrum-based method and the suggestion for supporting technology in form of a SOA/AOP layer towards earlier consideration of NFRs such as Privacy, Security, Trust and competing (N)FRs while building socio-technical systems such as di.me. AFFINE envisages involving experts or at least responsible(s) from each NFR category of relevance, e.g., legal and ethical concerns in order to ensure the right consideration from the beginning in the design and implementation of the respective system also at architectural level. TheAFFINE methodology is being now embraced by the company MT AG for the Integration Services business line.

[18] The solution's design process considers an attacker model and threat analysis.

[19] Santen began motivating his work by citing from Viega and McGraw (2001), who stated, "Bolting security onto an existing system is simply a bad idea. Security is not a feature you can add to a system at any time". He further argues, "the discipline of "Security Engineering" is far from mature today, and that, in practice, it still is not an integral part of the engineering processes for IT systems and software is based on the fact that security awareness results from reports on attacks – and not from the latest security feature that would make an application even more secure than it already was before."

implement technical support enabling protection of personal privacies would allow Europeans to continue valuing their right to privacy, even in the digital world, while allowing innovation in data brokerage and consensual, ethical commercialization of personal data.

Rising service-oriented broker platforms should consider law and ethics as core values during design phase, and in particular the concept of privacies, and existing OSNs should be required adopt these values if they wish to continue operating in the European market. Technically, such an adaptation could build upon the concepts of users having multiple context-specific digital identities, each of which serves for a particular purpose, and managing contextual information release. This could create a data exchange framework that respects law, ethics, and privacy without sacrificing commercial interest in data exploitation.

Di.me allows users to share personal information to other users and to other networks while providing the user with additional protection of their data, in particular by warning the user about the consequences of their actions if they have potentially unintended consequences. This protection is secure and allows the user to maintain control of his data, as the personal data is stored on the user's personal server – which could even be the user's laptop – and thus within the user's control. With a sizable percentage of the European population not being digitally literate, this approach could be important in enabling citizens to make informed decisions on exercising their right to protection and privacy of their personal data online.

Currently, the Directive is under review and may be replaced by a Regulation. One of the proposed changes is the strengthening of the principles of privacy-by-design by default and the promotion of data protection certification schemes. Moreover, standardisation initiatives will need to be promoted. Standardisation initiatives to ensure that social networking platform implementations are consistent with the revised data protection directive may be an interesting topic to investigate. di.me's APIs could contribute a basis for a privacy-oriented standardization intiative for cross-platform information brokerage of personal data. Further, the standardization mechanism could include a best-practice model for privacy-oriented design in social networking, to which di.me's approach could also serve as a foundational basis.

ACKNOWLEDGMENT

Even though just a few authors from the digital.me consortium contributed to this paper, it was surely not possible to compile such an article without the background efforts of all members ofthe di.me project. Special thanks are especially due to Fabian Hermann, Simon Thiel, Marcel Heupel, Ismael Rivera, BenediktWestermann, Lars Fischer as well as Philipp Schwarte.

REFERENCES

[1] J. Shim, M. Warkentin, J. Courtney, D. Power, R. Sharda, and C. Carlsson, "Past, present and future of decision support technology." Decision Support Systems, Vol. 33, Iss. 2, June 2002, pp. 111-126.

[2] L. Determann, "Social Media Privacy: A Dozen Myths and Facts," 2012 Stan. Tech. L. Rev. 7, pp. 1-14. [online] http://stlr.stanford.edu/pdf/determann-socialmediaprivacy.pdf

[3] "Conversion Rate Optimization." Blog run by Unbounce Marketing Solutions Inc. [online] http://unbounce.com/conversion-rate-optimization/

[4] I. Borges and D. Sinclair, "Media literacy, digital exclusion and older people." Brussels: AGE Platform Europe, December 2008. [online] http://www.age-platform.eu/images/stories/EN/pdf_AGE-media-A4-final-2.pdf

[5] S. Wheeler, "Digital literacies for engagement in emerging online cultures." Communication and Learning in the Digital Age, Barcelona: eLCRPS, Issue 5, pp. 14-25, November 2012.

[6] Articles 7 and 8, Charter of Fundamental Rights of the European Union. 2010/C 83/02.

[7] G. Santucci, "Privacy in the Digital Economy." The Privacy Surgeon, September 2013, p. 11 [online] http://www.privacysurgeon.org/blog/wp-content/uploads/2013/09/Privacy-in-the-Digital-Economy-final.pdf

[8] L. Lessig, Code and Other Laws of Cyberspace. New York: Basic Books, 1999.

[9] A. Cavoukian, "Privacy by Design … take the challenge." Toronto: Information and Privacy Commissioner of Ontario, 2009.

[10] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector.

[11] S. Thiel et al, "A requirements-driven approach towards decentralized social networks." Future Information Technology, Application, and Service Lecture Notes in Electrical Engineering. Vol. 164, Part 6, 2012. pp. 709-718.

[12] J. Stroh, Untitled post. Visual Metaphors community, 22 April 2013 [online] https://plus.google.com/100641053530204604051/posts/HDAVJBYBoSp

[13] O. Berg, "The Digital Workplace concretized". The Content Economy, 28 September 2012 [online] http://www.thecontenteconomy.com/2012/09/the-digital-workplace-concretized.html

[14] M. Bourimi, I. Rivera, M. Heupel, K. Cortis, S. Scerri, and S. Thiel, Simon, "Integrating multi-source user data to enhance privacy in social interaction," Proceedings of the 13th International Conference on Interacción Persona-Ordenado (INTERACCION 2012), art.51., New York: ACM, 2012, pp. 51-58.

[15] S. Thiel, F. Hermann, M. Heupel, and M. Bourimi, "Unlinkability Support in a Decentralised, Multiple-identity Social Network." To appear in the Proceedings of the Open Identity Summit 2013. Kloster Banz, Germany.

[16] M. Heupel et al, "Context-aware, trust-based access control for the digital.me userware," Proceedings of the 5th International Conference on New Technologies, Mobility and Security (NTMS) 2012.

[17] M. Bourimi, et al, "Towards transparent anonymity for user-controlled servers supporting collaborative scenarios," 9th International Conference on Information Technology: New Generations (ITNG), 2012. Pp. 102-108, April 2012.

[18] L. Fischer, M. Heupel, M. Bourimi, D. Kesdogan and R. Gimenez, "Enhancing Privacy in Collaborative Scenarios Utilising a Flexible Proxy Layer," Proceedings of the International Conference on Future Generation Communications 2012. London, UK.

[19] P. Schwarte et al, "Multilaterally secure communication anonymity in decentralized social networking," to appear in IEEE Xplore as part of the Proceedings of the 10th International Conference on Information Technology: New Generations (ITNG 2013).

[20] K. Cortis, S. Scerri, I. Rivera, and S. Handschuh, "Techniques for the Identification of Semantically-Equivalent Online Identities." 8194, LNCS, 2013.

[21] B. Gorriz and S. Thiel, "Package Structure," 17 October 2013. [online] https://github.com/dime-project/meta/wiki/Package-Structure

[22] European Parliament and the Council of the European Union, Directive 95/46/EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31 (23.11.1995).

[23] J. Dumortier, and C. Goemans. ‘Privacy protection and identity management’, in B. Blažičand W. Schneider (Eds.) Security and Privacy in Advanced Networking Technologies, Ios Press, 2004, p. 193.

[24] European Commission, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) COM (2012) 11 final – 2012/0011 (COD), 25 January 2012, commonly known as ‘draft Data Protection Regulation’.

[25] E. Kosta et al, "Requirements for Privacy Enhancing Tools." PRIME Consortium, 20 March 2008 [online] https://www.prime-project.eu/prime_products/reports/reqs/pub_del_D1.1.d_final.pdf

[26] J. Camenisch, R. Leenes, and D. Sommer (eds), Digital Privacy: Privacy and Identity Management for Europe (PRIME). Springer, 2011.

[27] J. Camenisch and E. Van Herreweghen, "Design and implementation of the idemix anonymous credential system," Proceedings of the 9th ACM conference on Computer and communications security (CCS 2002), New York: ACM, 2002, pp. 21-30.

[28] M. Heupel, M. Bourimi and D. Kesdogan, "Trust and Privacy in the di.me Userware," to appear in Kurosu, M. (ed) Human-Computer Interaction, Part III, HCII 2013. LNCS, vol. 8006. Heidelberg: Springer, 2013, pp. 29-38.

[29] M. Heupel, S. Scerri, M. Bourimi and D. Kesdogan, "Privacy-preserving concepts for supporting recommendations in decentralized OSNs," Proceedings of the 4th International Workshop on Modeling Social Media in conjunction with ACM Hypertext. Paris, 2013.

[30] A. Botzenhardt, H. Meth and A. Mädche, "Cross-functional Integration of Product Management and Product Design in Application Software Development: Exploration of Success Factors," Proceedings of the International Conference on Information Systems (ICIS) 2011. Paper 10.

[31] M. Bourimi, T. Barth, J. M. Haake, B. Ueberschär and D. Kesdogan, "AFFINE for Enforcing Earlier Consideration of NFRs and Human Factors when Building Socio-Technical Systems Following Agile Methodologies," Proceedings of the 3td Conference on Human-Centred Software Engineering (HCSE) 2010.