Project no. 610349

D-CENT

Decentralised Citizens ENgagement Technologies

Specific Targeted Research Project

Collective Awareness Platforms

D4.1 – State of the Art of social networking systems, identity ecosystem and social data stores

Version Number: 8
Lead beneficiary: ERCIM
Due Date: February 2014
Author(s): Harry Halpin
Editors and reviewers: Francesca Bria

Dissemination level:		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Approved by: Francesca Bria

Date: 28/2/14

This report is currently awaiting approval from the EC and cannot be not considered to be a final version.

Contents

1. Introduction

This report provides a broad overview of open standards that could be used by the D-CENT project in terms of decentralized and federated social networking, with a particular focus on our two use cases of democratic decision-making and social digital currencies. Current social networking platforms such as Facebook and Google+ are currently built almost entirely on proprietary technologies and cannot interoperate. Due to reasons ranging from data protection to the desire for autonomy by sovereign bodies such as community organizations and even nation-states, simply using these platforms "as is" is unacceptable to the D-CENT project. However, at the same point the D-CENT project must find a way to communicate with social networking sites such as Twitter and Facebook to take advantage of the "network effect" of the tremendous amounts of users on these sites. The ethos of the project is to develop software which is as open as possible. The use of open standards is distinct from the development of open source software.

The hypothesis put forward by this report is that the solution is the use of open standards that allow at least one open-source implementation.

First, we overview an argument for open standards rather than only open-source by learning from the failure of Diaspora, the most well-known alternative to centralized social platforms like Facebook. Then, we explain the difference between the various standards bodies and their licensing issues. In the rest of the report, we outline the landscape of standards in social networking and go through each part of the landscape in detail, ending with an analysis of socio-economic models for standards-based social networking and conclusions for possible next steps.

2. Open-Source is Not Enough: The Diaspora Project s Failure

Why open standards? Why not just open-source? For many programmers, using open-source software - or the more restrictive "free software" as defined by GPL licensing – is enough. However, the failure of the open-source Diaspora project should show that simply relying on open-source software is not enough to create a decentralized social networking platform. The Diaspora Project was sparked off by Eben Moglen in a speech entitled "Freedom in the Cloud," where Moglen critiqued how Facebook was destroying fundamental freedoms by centralizing human social interaction in a privately-run cloud without any accountability or ability for users to host their own data.[i] In essence, the paradigm of Richard Stallman and the classical "free software" model didn't make sense in the era of the cloud, as even free software components used by Facebook and Twitter were hidden behind servers and inaccessible to ordinary users.

Moglen countered that a personal server, called the FreedomBox, could be hosted by someone in their own house and so be under their physical control. The idea was that this device, which would appear similar to a television (cable) box, would eventually support privacy-aware applications and decentralized social software. However, at the time of Moglen's discussion none of the software for decentralized social networking had been created. This inspired NYU students Dan Grippi, Maxwell Salzberg, Raphael Sofaer and Ilya Zhitomirsky to create the Diaspora Project: a decentralized social networking platform. The project was launched with a large amount of fanfare due to an article in the New York Times entitled "Four Nerds and a Cry to Arms Against Facebook." (Dwyer 2010) The subsequent media uproar earned the students over $200,000 through Kickstarter. Cash in hand, the Diaspora team left NYU to work on the project full-time.

The central concept of Diaspora was that rather than using centralized social platforms, users would be able to set-up their own "pod" in a "peer-to-peer" network that let users directly share status updates and multimedia using encrypted messaging. The project began as a Ruby-on-Rails web application to be hosted on a FreebomBox or on a rented server in the cloud, though the Diaspora programmers also offered a central server called Diaspora* that could host users' pods. While the project gained thousands of supporters using the #joindiaspora hashtag, ultimately when the software was finally released in "alpha" in November 2011 it was difficult to use and suffered from a vast range of security and privacy related bugs. Worse, very few users actually ended up joining, so while Diaspora had over 200,000 users, most of the accounts remained inactive due to the bugs and usability issues. In summary, users who join Diaspora and look for their friends can't find them: they seem to be the only person in a large empty building. Lastly, rather than host their own pods, almost all users simply went to the centralized Diaspora* server at http://joindiaspora.com, and thus forsook actual decentralization in practice. The main feature that differentiated Disapora from Facebook, the idea of "aspects" that let users implement access control by dividing their friends into groups such as "work" and "family," was simply copied by Google+ before Diaspora was even launched and eventually even copied by Facebook.

Although the creators of Diaspora worked very hard to to fix security and usability problems, ultimately Diaspora ended up being an empty and more difficult-to-use clone of Facebook with an active user-base

of only open-source aficionados. The team began another Kickstarter campaign and began coding pods to allow information to come in from Facebook, Twitter, and Tumblr. However, it was too little, too late. Just as the project was propelled by the media, it was ultimately also destroyed by the media. A few days after the Wall Street Journal published an article called "What Happened to the Facebook Killer?" the "heart and soul" of the project Ilya Zhitomirsky committed suicide (Clayton, 2011). The rest of the original coders moved to try to create a new platform called http://makr.io for meme generation and the Diaspora codebase was "given to the community," which means that the original coders had effectively abandoned it. Since its being abandoned, there has been very little actual work on the codebase and user-base activity has been declining.

Why did Diaspora fail? One reason is that essentially the software failed to interoperate with the rest of the Web. While Diaspora claimed to be a decentralized peer-to-peer system, it simply locked users into an amateur "free software" project to replicate Facebook: users could not easily escape Diaspora as users and developers could not interoperate their own software with Diaspora. Rather than locking users into Facebook, where most users' friends actually were, users of Diaspora found themselves locked into Diaspora. The coders of Diaspora assumed that there was some inherent virtue in being free software (in particular, AGPL) that would give them success. Although open-source and free software does provide the ability for developers to easily fix bugs and update the software by crowd-sourcing from a larger community, for ordinary users the actual licensing of the software was for the most part irrelevant, as they had their "pods" hosted on the central Diaspora* server and thus downloading and installing new code was viewed more as a pain point than a critical feature. Also, simply using free software did not provide the necessary security and privacy features. While open-source and free software allows one the freedom to fork code, free software without good coding and integration does not magically allow a codebase to interoperate with the rest of the Web, particularly commercial providers but also even other open-source projects. In fact, there was a proliferation of open-source codebases such as identi.ca and Friendika, each with thousands of users despite their having much less media attention than Diaspora. As a fatal blow, most users already had their data and their friends on Twitter and Facebook, so there was no reason to use Diaspora as it did not provide any features not already provided by these commercial platforms and failed to interoperate with them.

3. What are Open Standards?

Open standards such as HTML and TCP/IP serve as the foundation of the Web and Internet. However, these standards follow a process that is quite different from the usual standardization process at either national-level bodies such as ANSI, regional bodies such as ETSI, or international bodies such as the ITU. These standardization bodies normally operate via formal processes that include democratic majority voting by representatives. In contrast, bodies such as the IETF and W3C involve consensus-making amongst a possibly open-ended group of members, where members are judged by the technical quality of their contributions, not what governing bodies they represent. In contrast to traditional standards bodies, this process is known as an open Multi-Stakeholder process. As the victory of the TCP/IP stack over the alternative ITU-backed network stack (OSI) demonstrated, it appears an open multi-stakeholder process is superior in creating the open standards. These multi-stakeholder standards bodies allow individual or institutional participation based on informality and consensus-driven decision making, with no official governmental or sovereign status: in the words of first Chair of the IAB David Clark: "We reject kings, presidents and voting. We believe in rough consensus and running code" (Clark, 1992, p 551).

The Internet achieves its stunning technical interoperability and equally global reach by bringing to the table a complex social network of interlocking and sometimes even inimical institutions ranging from open-source projects to companies such as Facebook who control the technical infrastructure. These institutions work together by deploying a number of standardized protocols that are "loosely connected" and respect the rather vaguely defined principles of Web and Internet architecture. These ubiquitous protocols, ranging from TCP/IP to HTML, were created and are maintained by a mixture of hackers, government bureaucrats or representatives of corporations that work via a small number of interlocking standards bodies such as the IETF or W3C. It is the responsibility of these standards bodies to create and defend, in the form of technical standards, the often implicit guiding principles of the Net and Web, such as net neutrality and the "end-to-end" principle, that are thought to have led the Internet to its astounding growth. The Internet Engineering Task Force (IETF) was created as an informal network of graduate students at the foundations of the Internet, who posted "Requests for Comments" (RFCs) for early Internet protocols such as TCP/IP and FTP. Frustrated with the large number of incompatible protocols and identification schemes produced by the IETF, Tim Berners-Lee had the vision of a 'universal information space' which he called the 'World Wide Web'. Tim Berners-Lee built the Web in his half-time at CERN as a quick prototype, with the Web being an application based on URIs (formerly known as URLs) that put a simple hypertext system on top of Internet. The core draft specifications (URL, HTML, HTTP) were sent to the IETF as "Experimental" specifications by Berners-Lee, despite having his original academic paper rejected from the 1991 ACM Hypertext Conference. However, when the IETF could not approve of a "Universal" or "Uniform" resource identifier scheme (URIs) quickly enough, and as large corporations started entering the IETF (with the possibility of abusing its informal process), Tim Berners-Lee began his own standards-body called the World Wide Web Consortium (W3C) to manage the growth of the Web. With offices in Europe, the United States, Japan, and even China as well as a paid neutral staff of over seventy employees, the W3C managed to rein-in corporate control of Web standards and has since fostered the development of the Web, including technologies such as HTML5. The W3C and IETF now work closely together, and with

ICANN serve as the core of the multi-stakeholder process of Internet governance described in the "OpenStand" principles.[ii]

One important part of the standardization process is not only the technical development of the standards, but the commitment to patents. While Europe has a reasonable patent policy, unfortunately in the United States there has been the excessive growth of software patents. With software patents, even open-source or free software can be claimed to implement a patented idea and thus be subject to licensing fees from the patent-holder. As software patents are a large business and there is a class of professional "patent troll" companies, it is important for any open standard to be free of patents so developers, especially open-source developers, can implement the standard without fear of licensing concerns. It is precisely the patent commitments by open standards bodies such as the W3C that allow commercial and open-source software to inter-operate: For example, W3C's licensing commitments to HTML5 allow both commercial closed-source browsers such as Internet Explorer and open-source browsers such as Mozilla to view the same web-page. This is done in the IETF by what has been called the "Note Well" agreement, namely that "In all matters of copyright and document procedures, the intent is to benefit the Internet community and the public at large, while respecting the legitimate rights of others".

The entire Note Well agreement[iii] explains this in much more detail in RFC 5378[iv] and RFC 4879.[v] However, the IETF does not guarantee royalty-free licensing via legally binding agreements. Given the high level of corporate competition in the Web, the W3C created itself as a membership consortium (although open to open-source developers, academics, government experts, and small companies via their "Invited Expert" process) so that these legal agreements can be made. These agreements essentially bind existing patents to the W3C, allowing the W3C to act as a "patent warchest" for all patents related to the Open Web, and then guaranteeing these patents are licensed royalty-free to developers everywhere. The W3C Royalty-Free patent policy is publicly available[vi] and the entire standardization process is described in the W3C Process Document.[vii] Companies are loath to violate the W3C's Royalty-Free patent agreement due to the fact that they would then lose royalty-free licensing on the Open Web Platform itself. The W3C in essence builds a patent "war-chest" around its standards, with commitments being made by its W3C's membership (thus the "consortium" in the W3C), which includes large patent-holding companies like IBM.

The social web is currently a very fragmented landscape, with a bewildering number of possible protocols and phrases. Due to the sheer size of this landscape, we will not review all technologies. We will define as a standard (in particular, an open standard) any standard that complies with the well-known W3C or IETF licensing policies, or that is currently in process, at the time of writing this report, into migrating into the W3C or IETF. We will not review other standards except in exceptional circumstances (as to be explained in the review), as we cannot guarantee they will be usable in a royalty-free manner by the developers in the D-CENT project. Thus, a large number of protocols, ranging from the Twister protocol[viii] or the Psyc protocol,[ix] will not be covered by this review. Given that these protocols in general have few users, this is not viewed as a disadvantage. The purpose of D-CENT is to allow users to build on a trusted and decentralized social web platform, so we will focus on standards that have some adoption as well as on well-known patent policies, which may serve as possible social substrata not just for the D-CENT project, but for the entire Web.

4. Topology of Social Networking Architectures

In general, we will consider social networking architectures as kinds of architectures on top of an application layer of services that a user accesses for sending natural language messages to other users on top of the network (IP) level. The issue of whether particular packets are delivered by alternative network architectures such as mesh networking is irrelevant. The application level is exemplified by email, status updates, chat, or even VoIP. The social graph is the set of users that a user sends messages to, usually on a repeating basis. The underlying applications that connect users to their social graph via messages, which may include routing user messages through client software and servers, provides a application-level abstract social network topology. Many different standard and software architectures can map to a social network architecture.

There are three broad types of topologies: peer-to-peer, where each sender can communicate autonomously to any other receiver without any mediating parties, centralized silos where every sender (client) must to go through a single organization (the silo) to communicate to a receiver (who must in turn receive the message via the same silo), and federated systems where every sender has to go through at least one server to communicate to a receiver, but the sender and receiver may be on different servers operated by distinct organizations. Examples of silos for messaging would include Facebook and Twitter, while federated social networking systems include Jabber chat and SMTP-based email. There are no widely deployed peer-to-peer messaging systems, but BitTorrent would be a good example of a widely deployed peer-to-peer system for file-sharing. While it has almost no users, Tribler would be an example of the use of a peer-to-peer architecture for social messaging (Pouwelse et al., 2008).

These social network topologies have a long history within network architecture, and with the differences in architecture first noticed by Internet pioneer Paul Baran of the RAND Institute in his seminal report (Baran, 1962). Baran classified three kinds of network topologies in his classic illustration:

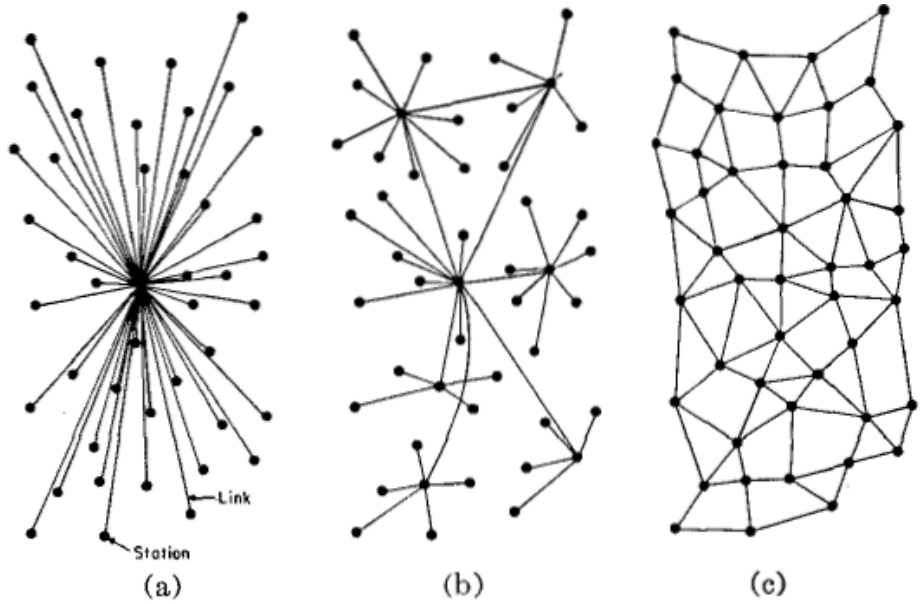

Figure 1 (a) CENTRALISED (B) DECENTRALIZED (C) DISTRIBUTED NETWORKS

In this illustration, Baran's fully distributed network is what we call the "peer-to-peer" network, while Baran's decentralized network maps to our "federated" network. Baran's distributed network maps to our definition of a peer-to-peer network. Since there is no commonly accepted way to refer to these different network architectures, there is a tremendous terminological confusion in this space, with terms such as decentralized, distributed, peer-to-peer and federated in general being mixed up. For example, the term "distributed" is usually conflated with "decentralized" and so a federated social network can often be called "distributed" in popular parlance. Also, often technologies are a hybrid of open standards and closed silos: The use of techniques such as the Facebook "Like" Button allows Facebook code to be embedded in any web-page, so that even centralized silos such as Facebook contain decentralized aspects. Lastly, the term "peer-to-peer" is perhaps the most confusing. We use the term "peer-to-peer" in a strict technical sense that distinguishes it from both federated and silo messaging systems, but we recognize that the term "peer-to-peer" (P2P) has become so popular due to a large amount of socio-economic work that almost any non-centralized system can claim to be "peer-to-peer" regardless of the underlying network topology. This terminological swamp can have harmful effects, as these three social network topologies lead to very different design decisions at the application level, each with their own trade-offs. In general, none of these social network topologies should be considered any "better" than any other in any absolute sense, but better than another topology in some more restricted sense in context of a particular use-case. For the given use-case of the D-CENT project, we will usually be assuming an underlying framework of the Web. As the Web is a federated system, with clients such as Web browsers communicating to servers, most of the standards we inspect here will be federated.

5. Security and Privacy Considerations

Although Snowden's revelations of how the NSA has taken advantage of the open nature of multi-stakeholder processes to abuse cryptographic primitives has been damaging to the credibility of organizations like the IETF and W3C, the organizations have responded by using their open process to determine how to best secure the Web against pervasive surveillance and undermining security by national governments, hosting workshops like the joint W3C/Internet Architecture Board workshop on Strengthening the Internet Against Pervasive Monitoring (STRINT).[x] Experts like Bruce Scheier, who did the technical analysis for the Guardian of the NSA leaks, have put forward that open standards in the IETF and W3C are the best way to address the privacy and security issues brought up the Snowden revelations.[xi]

Sadly enough, there has been little to no security and privacy analysis of the many federated and peer-to-peer protocols meant to replace silos such as Facebook. Revealing the extent of their amateur analysis of security, many enthusiastic developers such as those at Diaspora assumed that simply by moving to a decentralized or distributed protocol that they would improve the security and privacy of users. Far from it, for while central silos that are untrusted for various reasons are without a doubt a bad idea for the privacy and security of users, various federated and peer-to-peer schemes can be even more insecure and privacy-invasive. Every network topology makes certain design choices that privilege one security property over another. Although there may be no intrinsic trade-offs between these different security properties in regards to network topologies, so far the architecture of actual implementations belie structural bias towards certain security properties and against other properties.

Each of these architectures has a number of security properties: availability (ability for messages to be sent and received in a timely manner), usability (ease of use by user), authenticity (authentication of user identity), control (can the user move to a different server or client), anonymity (can the system be used without a personal identification), and unmappability (inability to detect the user's social graph). Since most standards are in the realm of federation, we will use some examples of the other architectures. For silos, we will focus on Twitter, Facebook and Google+. Peer-to-peer architectures generally include systems such as GNUNet[xii] and p2p social systems that build on it, such as SecuShare.[xiii] Federated systems systems include e-mail and Jabber chat. In the table below, we can see the trade-offs of each of these architectures in quite broad terms.[xiv] Do note that we do understand that reasonable security analysts may disagree over the exact values, and furthermore that we are describing only a class of network topologies rather than particular hypothetical systems or systems that do not have mass deployment. The terms “low” and “high” are actually relative, and relatively higher is not necessarily always good for users. For example, federated and peer-to-peer models have better authenticity than silo models, but usability problems make it so that their authenticity is often poor in practice.

Property	P2P	Silo	Federated
Availability	Low	High	Low
Usability	Low	High	High

Control	High	Low	High
Authenticity	High	Low	Low
Anonymity	High	Low	Low
Unmappability	Low	Low	Low

Unlike a silo like Facebook that runs continuously with high availability, in P2P networks an available low latency path to another peer cannot be guaranteed. This also holds for federated systems, although less so because the number of servers is relatively low so path latency tends to be higher and more robust. Also, P2P tend to be difficult to use, as they require that special software be installed to serve as P2P clients. Federated systems such as email and the Web also require special software to be installed, such as email clients and Web browsers, but as such software is generally already installed there is little problem. However, user control is normally low in silos, as users usually cannot retrieve their data and can by definition only communicate to other users of the silo unlike in peer-to-peer and federated systems. Authenticity tends to be high in peer-to-peer systems as peers can authenticate to other peers, but can remain anonymous insofar as they do not have to authenticate to any master server unlike in silos and federated systems. Yet in practice, peer-to-peer systems present usability for authentication, since most users do not use shared secrets or check key fingerprints. In contrast, federated systems often maintain a "whitelist" of other trusted servers and so can authenticate to each other, and can use various techniques to help overcome authenticity issues in practice. All topologies are failures at being unmappable.

Silos maintain mass amounts of personal data in a centralized form accessible to the owners of the silo and not the user, while federated systems at least fragment the personal data between multiple servers. However, even federated and P2P systems that allow anonymous usage reveal the social graph to outside adversaries rather easily via traffic analysis, which is perhaps the one saving virtue of silos: At least with a silo, an outside adversary can't determine one's friends. The information gained by mapping a social graph of any given user can usually reveal their identity even if a system allows users to join a communication channel without revealing their anonymity. For example, monitoring the patterns of communication in an IRC channel that allows anonymous identifiers can eventually reveal the identities of users of the IRC channel.

In general, for D-CENT, the main possible choice is between P2P and federation. However, peer-to-peer systems lack a trust model, which in turn allows attacks where the peer-to-peer system is infiltrated by malicious peers who can intercept traffic. This can be dealt with by using a friend-to-friend network topology, such that friends can share keys to encrypt content and then to route that encrypted content over "circuits"' with other friends in order to send messages without being intercepted by sybil attacks (Danezis et al., 2010). However, friend-to-friend topologies are susceptible to traffic analysis as they reveal the social graph of the user. P2P designs such as Drac overcome this problem to some extent by hiding the friend-to-friend topology in a larger network topology. However, the first `hop' in the circuit between the sender of the message and their first `friend' in the friend-to-friend topology circuit is still exposed: One of the assumptions of Drac that figured into exposing the `first hop' was that the local social graph of the user was already known. For example, an activist in Damascus would

already have their entire `local' social network known by their adversary, but would wish to disguise any messages sent to activists in remote locations such as another activist group in Homs or a support group in Paris. Yet in real-world use-cases, the direct friendship network is not necessarily local or already under observation, and thus defending this part of the social graph is important.

Second, in order for messages to be sent efficiently, nodes must be connected to each other via a small number of `hops' between peers, and highly well-connected nodes (supernodes) are the solution to this problem of availability. Interestingly enough, peer-to-peer systems tend to naturally evolve into centralized network topologies around supernodes with vastly unequal numbers of connections in a `power-law' distribution. Peer-to-peer systems like BitTorrent and Skype often have these supernodes artificially created by the designers of the networks in order to encourage efficiency (Guha, 2006). If these supernodes are actually hostile, they are capable of observing a large amount of traffic. The traffic between non-supernodes is `thin', i.e. infrequent and routed possibly through multiple non-supernodes, which makes anonymity and unmappibility difficult as there is simply less regular traffic to hide amongst. Given these disadvantages, it is unclear why peer-to-peer architectures are often preferred, and one suspects a certain ideological preference could be being projected into technological design choices. Given that peer-to-peer systems naturally evolve "servers" in the form of supernodes, why not use servers?

Instead of focusing on pure peer-to-peer architecture designed for a hostile network, a client-server approach assuming a trusted server can be both more secure and more realistic for deployment as it would not have the issues around NATs and installing new client software that a non-Web based approach would have. In our approach for D-CENT, federation makes sense so that the architecture is such that there is an initial connection by a client to a trusted server, and then the network is routed between friendly servers in the federation as the servers have a "friend-to-friend" network topology. Attackers can be detected as they must consciously join the federation and obey the normative rules of the federation. Although describing the normative rules of joining the federation of trusted servers is outside the scope of this deliverable, work on governance and the commons would be useful here. Furthermore, in general neither peer-to-peer nor federated network topologies have been adequately studied in terms of privacy and security, and neither are a magic pill. While there have been national-level efforts such as the SPION project in Belgium, these projects have studied the rather egregious privacy failures of large commercial systems such as Facebook.[xv] Their main results have been "toy" systems to send encrypted messages over Facebook or to try to provide guarantees of context in sharing, and they have not looked at entirely new non-silo social networking, identity or data-store architectures. Given the immature level of research in this area, at this point we recommend that any standards in this space have adequate security and privacy reviews, as will be done via W3C process.

Other EC projects have results that could vastly improve the anonymity and authentication aspects of existing social networking systems. There have been a number of relevant EC projects in this space, in particular the EC project PRIMELIFE and the related Network of Excellence ABC4TRUST.[xvi] These networks have been focussed on research and implementation of cutting-edge "zero-knowledge proof" (attribute based credential) systems for authentication of users that allow the user to both prove who they are without necessarily revealing any of their personal data, or even an identifier such as a name or user-name. These technologies have much promise and could be used by a federated, peer-to-peer, or even silo-based social networking system. While it is beyond our scope to explain the complex cryptographic algorithms behind zero-knowledge proofs (Fiege et al. 1987), the systems do have

promise, as shown by the purchase by Microsoft of the zero-knowledge proof authentication system U-Prove.[xvii] Luckily, EU funding has produced an open-source version of zero-knowledge proof authentication systems called IDEMix, with the lead being taken by IBM Research (Camenisch and Van Herreweghen, 2002).[xviii] Although such a system is technically open-source, IBM Research still holds patents on the underlying technology. In conjunction with the W3C, in order for this technology to be used by D-CENT it should be standardized and released under a royalty-free patent policy. This is challenging but is where W3C efforts towards standardization will happen in the project.

Given the advantages and disadvantages of every network topology and solution it engenders, technical solutions will not be enough to guarantee the enforcement of privacy and so legal-based privacy enforcement techniques should be encouraged in order to help the adoption of D-CENT. Thus, the D-CENT project should take advantage of moves by the European Union towards a single digital market, with a particular eye towards legislation to defend the fundamental rights of users as well as build European trusted services. In this regard, although the zero-knowledge-proof attribute credentials were removed from the amendments to the eSignature directive, the current updated directive still includes better liabilities for breaches and breach notification which should encourage higher levels of privacy and security in alternative social networking systems, as well as improved authentication standards. In this regard, recent moves to improve and unify a European-wide Data Protection act are tremendously important. This act, although counter-balanced in terms of anonymity and unmappability by the Data Retention Act, still has strong requirements for data portability based on the kind of open standards that we inspect below. Lastly, as the D-CENT project will use open standards for social data, data produced by this project should be in harmony with current and future work in Europe on improving open data standards.

6. Identity Eco-systems

Identity is the connection between descriptive data and a human or social institution, and as such essentially serves as the "digital name" of some entity. As the Web permeates more and more of everyday life, the lines between one's digital identity and one's offline identity are increasingly blurred and indistinguishable. Identity systems have long been extended beyond simple natural language names such as "Tim Berners-Lee." For example, Tim Berners-Lee may have a phone number, which with an internationalized calling code would be a USA phone number of 10 digits. These digits, such as +1.617.253.2613, are not human readable or connected to the Internet in an obvious way. However, with the advent of the Internet, a number of new identification schemes came into being, such as email like timbl@w3.org, or even Facebook accounts like "Tim Berners-Lee" that map to a URI controlled by Facebook: https://www.facebook.com/tim.bernerslee. Interestingly enough, while one's natural language "proper name" is in general registered and controlled by the government, identifiers ranging from phone numbers to e-mail to Facebook accounts tend to be controlled by either non-profit organizations, or more typically private corporations.

This proliferation of identifiers that have no standard ways of interoperating or even connecting has led to the vision of an identity eco-system, where the various identities and relevant data of persons and organizations could be connected together in order to enable new services and more efficient transactions. After all, how does one connect one's Facebook account to one's email, to one's phone and to one's national identity card? This problem is thorny and complex, and no easy solution has so far emerged. There have been various different calls for how this is to be done. The first step is to choose what identifier gets used to connect the other identifiers together, which in turn determines who controls the identifier.

The earliest is a rather individualistic approach, called the user-centric identity eco-system, as put forward in the first paper to really broach the question of user-controlled identity and personal data, "The Augmented Social Network: Building identity and trust into the next-generation Internet" which proposed to "build identity and trust into the architecture of the Internet, in the public interest, in order to facilitate introductions between people who share affinities or complementary capabilities across social networks" in order to create a "a form of online citizenship for the Information Age" (Jordan et al., 2003). Although the paper was ambitious in scope and wide-ranging in a vision for "revitalizing" democracy, the concept was not taken to a standards body. Instead, an entrepreneur called Drummond Reed of a company called InterMinds created a new kind of identifier called XRIs (Extensible Resource Identifier) (Reed and McAlpin, 2005) . An "=" meant a person and an "@" meant a company, such that a person such as Jeff Hodges would have the XRI =JeffH while a company like Paypal would be @Paypal. However, what Reed hoped to do was to replace user-centric identifiers with a for-profit monopoly on identity controlled by himself.[xix] When he claimed there were patents on XRIs on a public mailing list, Tim Berners-Lee called for them to be rejected, and the W3C intervened so that the proposed XRI standard was rejected from the OASIS standards body (a standards body for Web standards that includes non-royalty free standards, unlike the W3C).[xx] Nonetheless, Reed persisted for years in trying to instil XRIs in various decentralized social networking technologies, including OpenID and WebFinger. So, while the vision of an identity scheme controlled by users remains very compelling, in the United States this vision seems to simply lead to another group of start-ups wanting to control user identities,

as exemplified by groups such as the Personal Data Ecosystem Consortium.[xxi] Of course, this very North American approach begs the question why a small for-profit start-up would be more trusted with a user's identity than a large corporation such as Googlefd or Facebook.

However, the lack of robust identities within the United States has continued to be a problem, particularly as there is no nationally mandated identity scheme. The National Strategy for Trusted Identities in Cyberspace (NSTIC) is a 2011 U.S. Government-sponsored attempt to start a self-regulated identity eco-system within the United States.[xxii] Without a national-level identity system, the United States government had trouble maintaining databases on its citizens. Since large companies such as Verizon maintained accurate records of people's addresses and companies such as Google controlled digital identities, the original thrust of NSTIC was that both the government and large companies such as banks could save massively and connect disparate identities by outsourcing their databases of identities and relevant personal data (such as addresses) to private corporations. However, rather than enable user-centric or even small enterprises to compete via open standards produced by the W3C and IETF, the NSTIC effort became controlled by large players such as Paypal and used "homegrown" standards such as OpenID Connect. So despite promises to produce vast savings and improved authentication, the results in unifying identity for ordinary users of NSTIC were non-existent, although a vibrant backend in data verification via trust frameworks slowly began. Within Europe, identity has traditionally been more trusted to be managed by national-level governments than by private corporations or users, and while countries such as Estonia already feature large-scale eID systems, current EC work in this space such as the STORK Project[xxiii] have only just begun to look at issues of user-control and privacy. Furthermore, their approaches attempt to map identities on the national-level, but do not take into account in any detail the proliferation of digital identities.

The question of what precise digital identifier to use still remains open. Tim Berners-Lee supports the use of URIs (Uniform Resource Identifiers, previously known as URLs or "Uniform Resource Locators") as identifiers not just for web-pages, but for all sorts of things. This universal scope of Berners-Lee's ambition is even in the IETF URI specification which states that a URI identifies a "resource" where, "A resource can be anything that has an identity. Familiar examples include an electronic document, an image, a service (e.g., 'today's weather report for Los Angeles'), and a collection of other resources. Not all resources are network 'retrievable'; e.g. human beings, corporations and bound books in a library can also be considered resources." (Berners-Lee et al., 1998). For Tim Berners-Lee, his URI would be http://www.w3.org/People/Berners-Lee/card#i. This particular URI is not his homepage, which is just http://www.w3.org/People/Berners-Lee/, but a fragment identifier attached to a page for his VCard that allows people to distinguish a URI for "Tim Berners-Lee the person" from "Tim Berners-Lee's Web-page." However unlike in XRIs, there is no standard way to distinguish URIs for people or organizations from those of web-pages. However, URIs allow one to leverage the considerable infrastructure of the Web to solve the Discovery Problem, the well-known problem of how to discover and find users and the capabilities a user may support in a user-centric identity eco-system. In particular one can deploy the "follow-your-nose" algorithm to retrieve data from a URI using a HTTP GET request in order to discover more data. One possibility is the IETF draft standard for subdirectories like .host-meta and more generic .well-known subdirectories from any URI (Hammer-Lahav and Cook, 2011), or even using HTTP Link Headers that can include custom links to application-specific content using the IETF standard for Link Registries (Nottingham, 2010).

Unfortunately, although URIs as identifiers for people and other things beyond web-pages is the foundation of the Semantic Web technology stack, very few users identify with a URI as their primary identifier and find them hard to remember. Due to this, technologies that have been based on forcing the use of URIs as identifiers such as early versions of OpenID or WebID have failed to gain much traction. Far superior in terms of memorability are email addresses such as timbl@w3.org. E-mail addresses are very personal, so that users identify with them naturally, and they are usually easy to remember. They are also, unlike URIs, associated with a concrete set of embodied actions, i.e. checking and reading email inboxes. The first standard proposal to replace URIs as identifiers with email addresses was WebFinger, an IETF standard that re-designed the classic finger UNIX protocol to use e-mail addresses to solve the discovery problem (Jones et al., 2013). While often viewed as competing with URIs as a primary identifier, email addresses can easily be transformed to URIs either in some custom way or via a standard URI scheme such as the acct URI scheme. Thus, timbl@w3.org transforms into the URI acct://timbl@w3.org. This approach also works with custom Twitter or Facebook handles, so that the Twitter handle @timbl would transform into acct://timbl@twitter.com. WebFinger has seen some deployment across federated social codebases and even support from Google and Microsoft at times, so in general it seems that having users use and remember email addresses is superior to URIs, even though there are edge-cases such as shared e-mail addresses that are not addressed by this technique.

However, the main weakness of both URIs and email addresses is that they depend on the domain name system. In other words, users do not actually control their own names, but the owner of the domain name does. So for timbl@w3.org, the W3C controls the domain name. If Tim Berners-Lee left the W3C, he could lose both his URI and his email address. Some services, such as Yahoo!, recycle old email addresses, so that the identifier tim@yahoo.com may be given to multiple users over time. While the W3C may seem to be a trusted host, in the case of services such as Facebook, Twitter, and Google, ultimately for Facebook names, Twitter handles, and Gmail addresses, the identity of the user is completely controlled by the corporation and the user has no rights over their identity, which is perhaps a worse situation than that faced by identities assigned by nation-states. Furthermore, even corporations do not own domain names indefinitely, but lease them from domain registrars who ultimately lease them from ICANN,[xxiv] which has the IANA (Internet Assigned Names and Numbers Authority)[xxv] function to distribute domain names on lease from the U.S. Department of Commerce. Furthermore, the domain name system is centralized, with the root domain name servers being in the United States, with only two (one in the UK and one in Sweden) outside of the United States. However, this situation may change as the NSA revelations have diminished trust in the United States and so the upcoming meetings in Brazil will likely try to move control of the IANA function into a more global body than the United States.[xxvi]

On the other end of the spectrum, attempts to create a fully decentralized identifier system based on distributed hash-tables have flourished, but none have been standardized. For example, a Bitcoin-like mechanism to replace domain names called Namecoin[xxvii] has seen some success. Another popular approach is telehash,[xxviii] which sends small data packets (in JSON) using UDP (the protocol used by Skype and others) with a Bittorrent P2P file distribution mechanism. However, such fully peer-to-peer approaches have not been tested on a large scale and one loses the ability to rely on the Web infrastructure, and such approaches have also not yet been through the IETF or W3C standards process. Worse for ordinary users they are forced to use an indecipherable cryptographic hash instead

of a human-memorable identifier for their identity. While Tim Berners-Lee may not think timbl@w3.org is a great identifier, he would surely balk at using "13630378882" as his identifier.

Note that identifiers do not have to be identified with a single person's real name, but can also be identified with pseudonyms or with "one-time" temporary identities for anonymity. A single domain name, such as http://www.example.org, can easily produce pseudonymous identities such as http://www.example.org/batman or infinite one-time identities such as http://www.example.org/anon13630378882. Such a system can easily be repeated for email addresses, such as batman@example.org or anon1363037882@example.org. The creation of multiple identities with different characteristics is often considered a hard requirement, particularly for users that have legitimate concerns over stalking and free-speech, as well as to encourage creativity. The maintenance of multiple identities is usually called personae in identity ecosystems, and the D-CENT project will recognize multiple identifiers without "real name" restrictions such as those employed increasingly by Facebook and Google.

The main problem facing the use of email addresses as identifiers is actually that of the domain name system. To be purely technical, identities on the Internet and Web are not a commons: Right now domain names, like IP addresses, are a common pool that is privatized via ICANN and then domain name registrars, and so is in the long-term unsuitable as an identity-system. However, a new identity layer for the Internet and Web could be created that requires a kind of legal agreement for creating names and then permanently removing them from circulation from the domain name system. This could be done via a request from IETF or W3C to IANA, although more research should be done on commons-based governance and other models for running these kinds of identifiers in order to determine how they would differ from the domain name system. However, we imagine such an identity layer could be implemented in parallel to the current domain name system with simply a new set of socio-technical rules and requirements, and thus these new kind of identifiers would be the same technically as email addresses as URIs, just with a new kind of domain name attached to the end, such as timbl@commons.example.eu.

7. Personal and Social Data Stores

The main use of an identity eco-system is to enable the use of personal and social data. Currently, centralized silo social networking systems such as Facebook, Google+, and Twitter mostly track low-quality social data, such as names and the social graphs. However, there have been moves towards enforcing higher quality and verified personal data, such as the use of a "real name" policy in Google+ and attempts by Google+ and Facebook to link phone numbers as well as geolocation to identities in their proprietary silos. However, high-value data such as credit histories and medical records are to a large extent still the reserve of traditional institutions such as banks and hospitals. The thesis put forward by the World Economic Forum in reports such as "Personal Data: A New Asset Class" (World Economic Forum, 2011) is that this high-quality personal data that is currently "locked" in traditional institutions could serve as a valuable input into data-driven innovation, which in a neoliberal framework could enabling a whole new realm of efficient enterprises and start-ups. In a more European vein, such high-value data could also enable community-based social innovation.

In general, the thesis is that users should control their own data via personal data stores, also called "personal data lockers" or "personal data zones." These personal data stores consist of attributes, such as full name, phone number, bank balance and medical attributes. Various systems can be used to double-check these attributes by various means including machine-learning and background checks, and so create verified attributes. By controlling their own data, users could then engage in contracts that would enable powerful services in exchange for their data, establishing trust frameworks via algorithmically-backed and legally-binding agreements. Some people like Jaron Lanier even imagine a form of micropayments for the use of personal data (Lanier, 2013). However, the imagined data liquidity to be enabled by these personal data eco-systems have yet to be realized for ordinary users.

Instead, efforts like the National Strategy for Trusted Identities in Cyberspace have produced a number of trust frameworks for verifying attributes by traditional institutions such as banks and corporations, but there is no user-control of data or the verification procedure. These trust frameworks that have been produced by NSTIC are exemplified by the certification procedures enabled by groups such as the OpenID Connect-based Open Identity Exchange[xxix] and the Kantara Initiative.[xxx] However, these "open" trust frameworks are likely dwarfed in size by the vast amount of verification procedures being produced in the background over user's data by a combination of government agencies and private corporations. The long-term strategy of trust frameworks and verified attributes may even lead to conflict between some of the established banking institutions and identity silos such as Facebook, Google, and Paypal whose payment systems may eventually be viewed as a threat. For an example of verified attributes in action, the reason why Paypal transfers take longer than normal credit card transfers is the longer-time needed to take into account verification of attributes on the low-cost payment network used by Paypal, which can done with higher accuracy and with more managed risk using banking verification networks.

Yet in general, ordinary users, governments, and communities in Europe have been left out of this evolution of trust frameworks and personal data stores. There are two reasons: namely, most in action trust frameworks do not rely on open IETF or W3C standards and are without open source implementations, so it is unclear how new actors can seize control and create their own data and trust

networks. Second, on a conceptual level the real value of data is actually social, and the traditional actors such as banks and marketing agencies are more interested in individual humans than in social groups. Thus, personal data tends to be limited to profiles, collections of attributes about individuals as given by traditional Facebook or Google+ profiles, rather than progressing towards a more full notion of socially-produced data. However, the recent focus on extending profile data to include the social graph of the user as well as the dynamic interactions of activities (comments, posting content, "like" usage) shows that a more full notion of social data is slowly gaining traction. This leads to the possibility of not just individual personal data stores, but social data stores for data produced by communities that are in turn governed by these very communities. However, we do not have many examples of working software for social data stores, much less working models of governance.

To return to the technical level, even the most complex of governance frameworks for trust frameworks and social data-stores must ultimately rely on the transfer of data. This requires that data be automatically created, parsed and processed using a standardized data format. Evolving from the text-only ASCII character set to a fully international Unicode data-set has proven to be very useful for natural language, but social and personal data is usually highly structured. Early attempts to produce complex data structures used ASN.1 but ASN.1 is notoriously difficult to parse (ITU, 1984). This problem is particularly important insofar as parsing ambiguities can lead to actual security exploits in important ASN.1-based formats such as that used in client and server certificates (Kaminksy, 2010). There was movement away from pure text formats to structured formats for documents like HTML by the generalized markup language XML (Extensible Markup Language), a W3C standard (Bray et al., 1998). Although it was well-suited for tree-structured data, XML did not use URIs for identifiers and proved to not easily map to common data structures. The W3C created the RDF (Resource Description Framework) standard for using URIs as identifiers in graph-based data (Klyne and Carroll, 2004). As URIs are used in this data format, RDF-based data can be easily linked and merged by using URIs, and so deployments of RDF are called Linked Data (Berners-Lee, 2006). However, most data structures such as hash-tables (associative arrays) and ordered lists are not naturally suited for use in the graph-based RDF or tree-based XML language, so a very simple and easy to parse language called JSON (JavaScript Object Notation) has been gaining in popularity (Crockford, 2013). As RDF is an abstract semantic model that can be serialized in a wide variety of syntaxes, RDF has recently moved beyond the universally-detested RDF/XML format to JSON-LD (JSON Linked Data) that makes RDF compatible with JSON.

7.1 UCard, FOAF, and Personal Contacts

VCard is the first IETF standard format for personal data, based on data typically found on a business card, such as phone numbers and full names, and was given a syntax in plain ASCII text (Perreault and Resnick, 2011). Each vCard is attached to a single individual, and coming along as vCard did before the growth of social networking, vCard lacks social graph information.

This lack was taken up by two different proposed standards in the realm of RDF and XML. For RDF, the FOAF (Friend-of-a-Friend) format focused on describing social relationships as a graph, and although it was greeted by initial support, it ultimately failed to achieve much adoption (Brickley and Miller, 2014). Although it has the advantage of using URIs as clear identifiers, this was due to eventual divergence from more popular vCard and PortableContacts specifications, and was further held back by lack of developer interest in RDF and lack of tooling, although FOAF remained popular in academic circles.

Large corporations instead focused on supporting XML and JSON formats, in particular as given by the simple and data-driven PortableContacts specification that was designed to easily map to popular profiles such as the Facebook profile page (Smarr, 2008). Not only a format, it also contained an API for easy access in applications as well as the use of OAuth for authorized access to personal data, and was compatible with other industry efforts such as the OpenSocial API.

Today, all standards are converging around the vCard 4.0 effort at the IETF, including FOAF and PortableContacts. Recently, the VCard 4.0 IETF standard allowed properties about groups and organizations as well as social graph information [VCARD4]. VCard 4.0, influenced by the popular PortableContacts XML format for the social graph, also produced an XML format. Currently a mapping also exists to RDF, allowing vCard to be used in JSON. Given the widespread use of vCard in mail clients and mobile phone address books, it is expected that vCard may remain the de-facto personal data standard. Still, considerable work needs to be done to allow vCard to be used as the basis for social datastores.

8. Authentication Standards

The act of claiming and then using an identity can often be broken down into two distinct elements: authentication and authorization. The first step, authentication, is when some proof of identity is offered. A particular poof of identity is called a credential. Credentials can take forms as diverse as the display of an identity card to the proof of possession of a secret, such as a password or private key material. Identity providers require authentication to access services and resources, and may also associate (possibly verified) attributes with an identity. Note that authentication does not necessarily reveal any identifying characteristics and so may keep the authenticating entity anonymous, as using techniques such as "zero-knowledge proofs" allow a user to authenticate without revealing their identity such as a legal name or other attributes to the identity provider. Of course, different kinds of identity providers or even the same identity provider may host different personae, and different levels of security may require different kinds of credentials. An identity provider or user should be able to revoke an identity due to the possibility of compromise or some other valid reason.

While identity cards are currently the predominant way to prove identity in the "offline" world, on the Internet the combination of a username with a password is simply the de-facto method of authenticating users on the Web. Usernames and passwords are an easy to use technology, as both the username and password can be human memorable. Current best practice on the web uses an HTML form for entering a username-password combination, which returns a cookie (a small identifying piece of code kept in the browser) that tracks a user's session state (i.e. "logged-in" and any other information, such as purchase information and viewing history). Between every session, users have to re-enter their password. While this is currently a very popular technique, passwords are often a pain point for users who have trouble remembering their password and for developers who lack consistent and simple Javascript libraries for user-name/password authentication.

Worse, passwords suffer from a number of security issues that are increasingly making them untenable. Primarily, the very fact of making them human memorable makes them insecure, as usernames are often public and passwords, by virtue of being human-memorable, are often easily captured by phishing techniques or even automatically guessed by techniques such as rainbow tables. As users have cognitive difficulty remembering differing passwords across multiple web-sites, users often re-use the same username and password combination for all their accounts, so that when one account is hacked their entire online life can be hacked: The more sites a user visits with the same password, the larger the potential damage of a hack. Worse, often passwords are often simply insecure on the server-side database, are stored using simple hashing techniques (usually MD5) that can be cracked via online services such as Hash-cracker.com. Thus, when a server is compromised – as happened with the Target data breach – the passwords and data of millions of users can be compromised. Even cookies are not very safe, for without enforcing TLS correctly, the cookie that stores session state is easily swiped by a man-in-the-middle attack by an outside attacker.[xxxi] IETF standards such as HTTPAuth do not use encryption or use broken encryption (Franks et al., 1999), but at least have the advantage of having the user enter their password in a form based in the browser rather than inside the Web, which is less vulnerable to attacks such as cross-side scripting attacks.

Particularly in sectors like government sectors and enterprise, there has been a focus on moving away from only username-passwords and instead using some higher-security framework. Recently, as more and more of users' valuable personal and social data find themselves within identity eco-systems, there has been work on increasing the security and usability of user-facing authentication. In cryptographic terms, a username-password is a symmetric shared secret, so that this username-password combination is by itself enough for authentication. In general, higher-security measures involve moving from shared secrets to asymmetric public-private key encryption, where a user has to possess some secret private key material as a credential for their identity. The advantage of this approach is that even the server (who possesses the data) cannot access the data without both the public and private key, and the private key does not have to leave the user's browser, device, or smartcard. Perhaps the most simple technique that combines an easy-to-use password with private key material in a patent-free method is the Secure Remote Password Protocol (SRP) which in essence creates a shared private key from a user's password and can use that to authenticate a user and then encrypt the session state without reliance on third parties or complex certificate infrastructures (Wu, 2000). However, currently HTTPAuth does not support SRP and so is not easily done in user-facing browsers. The W3C is encouraging the IETF to upgrade HTTPAuth to use SRP rather than its current insecure techniques.

Within government and enterprise environments, the SAML (Security Assertion Markup Language) standard is currently the most popular. SAML is an OASIS standard for the exchange of authentication that uses an XML-based data format featuring rich metadata, ranging from the subject making an assertion of authentication or attribution, the time of the attribute, any conditions, and the resource to be accessed.[xxxii] A trusted third-party identity provider verifies these assertions and uses them to determine authentication using a number of different options, ranging from simple passwords to private-key credentials, and often can work in multiple federations. Examples of its successful use in federated environments include the Eduroam network of universities. However, for user-facing scenarios SAML is normally viewed as too complex and mostly can be used in environments where restrictions can be put on computers at work and employees can be forced to download certain software or even carry smartcards for authentication.

8.1 WebID

One simple private key solution is the informal WebID specification that uses the private-key material in client certificates produced by TLS (HTTPS) for authentication (Story and Corlosquet, 2011). WebID is maintained by the WebID W3C Community Group. However, the WebID Community Group is not an official W3C Working Group, and so lacks some of the stronger patent protection of the Royalty Free Patent Policy and W3C and industry endorsement. To authenticate a user requesting a resource, the identity provider controlling the resource needs to request a self-signed X.509 certificate (ITU standard for private key certificate formats in ASN.1 used in TLS) from the client. Inside the client certificate there is public key material as well as a field which contains a URI for the identity of the user (which is technically the "WebID", with the larger protocol called WebID/TLS). Using TLS, the browser confirms they know the private key matching the public-key in the certificate as well as the public key in the URI profile. The advantage of this protocol is that the user does not have to remember a password or username and it uses private-key material that relies on the well-deployed TLS infrastructure. However, the approach suffers from a number of disadvantages that have limited its deployment to Semantic Web hobbyists and prevented any real uptake by industry (as thus the W3C), as client certificates are currently not supported by browser user-interfaces and do not work cross-browser, much less cross-

device. The WebID protocol also requires a user to identify and control a URI, which many users do not.

8.2 BrowserID

Another high security approach backed by Mozilla is called BrowserID, recently re-branded as Mozilla Persona which is to be used as the primary identity and authentication system by Mozilla.[xxxiii] Rather than relying on identity providers such as Google authenticating and shipping personal data server-to-server, using BrowserID all authentication and attributes are sent via the browser. However, rather than use a client certificate like WebID, BrowserID uses a "verified'" email as the primary identifier and the possession of the browser and the e-mail provider with private key material as the primary credential for authentication. Thus, a user needs only to register an email as their identifier and then should be able to authenticate without an additional password, and private key material controlled by the browser can be used to authenticate with a higher-level of security assurance. As the entire authentication flow happens in Javascript, an easy-to-use developer API is provided by Mozilla and the user-interface can be very friendly, unlike in WebID or HTTPAuth, and it can work cross-device by relying on the browser and Javascript to synchronize the key material required. However, BrowserID has also seen almost no adoption outside Mozilla. First, it requires the co-operation of email providers for real security assurances, and few email providers have co-operated by registering key material for their users. Second, it only works when the user is online with their browser, which is both a possible privacy benefit but also makes it difficult to use for many use-cases. Third, as the protocol was never shipped to a standards body like W3C, many people believe that BrowserID only works with Mozilla.

8.3 Authorization Standards

The second step of identification is authorization, where after there has been a successful authentication of a user, the user can then authorize the transfer of attributes between services. The service with the attributes is typically called an identity provider, in particular if they also offer authentication services, but also called simply an "attribute provider" if they only offer (usually verified) attributes. A relying party is the service that wants attributes. The typical user-case is that a user wishes to log-in to a new service and wants their profile – including name and picture – to show up in the new service, and so authorizes an existing identity provider such as Facebook to transfer those attributes to the relying party. If not already logged into Facebook, this is typically done via being redirected to Facebook and then authenticating to Facebook via a username-password using their proprietary Facebook Connect flow, and then asking the user to explicitly approve the relying party access attributes stored on Facebook. After the user accepts the transfer of personal data, they are redirected back to the now personalized new site. This approach is also being copied by other major identity providers such as Google and even Twitter. Besides the usual problems with authentication already mentioned, there are security disadvantages to this approach. First, getting users accustomed to this kind of flow is "phishing heaven" as the redirect to Facebook and other providers can easily be faked, as most users would not notice if the URI was not the real URI as long as the login forms looked correct. From a privacy standpoint, the identity provider observes all personal data transactions with any relying party and so builds a map of the user's use of other services. Worse, there is nothing technically preventing an identity provider from doing personal data transactions without the user's consent.

8.4 OAuth

When authorization first started, users often had to give their username-password combinations to relying parties, which is a terrible security flaw as it allows the relying party unlimited access to personal data on the identity provider. The IETF standard OAuth (Open Authorization) 2.0 lets users share attributes on identity providers with relying parties without having to give the relying party these credentials (Hardt, 2012). OAuth instead enables authorization to happen via a shared secret with precise time and access permissions, but still makes the relying party redirect the user to consent to the personal data transfer on the identity provider in order to produce the shared secret (OAuth token). The OAuth 1.0 approach even works securely over ordinary HTTP requests, as the client generates a signature that contains unique information to the shared secret, preventing cookie-snatching type attacks (Hammer-Lahav, 2010). However, the standard has proven to be vulnerable to session-fixation attacks by saving old requests for authorizations and fooling a user into re-consenting, but this was fixed by registering relying parties to identity providers. Also, timing attacks were shown to work, but new work on fixing the signatures (i.e. using constant-time) fixed these attacks. OAuth 2.0 attempts to simplify OAuth 1.0 by forcing TLS (HTTPS) encryption (Hardt, 2012). Further work with client-generated signatures, possibly produced by the W3C Web Cryptography API, could force end-to-end encryption of the authorization, preventing many future attacks (Sleevi, 2014). However, OAuth is more of a generic framework and still is normally used for user-name password authentication via redirection, and thus has the same security problems mentioned earlier. OAuth is currently supported by large providers such as Google and Twitter.

8.5 OpenID Connect

While the original OpenID 1.0 specification used the proprietary XRI work and was based on a difficult-to-use XML stack, newer versions such as OpenID Connect essentially are profiles of OAuth 1.0 for identity-based attributes.[xxxiv] Thus, the essential security properties and information flow of OpenID is the same as OAuth. Large identity providers such as Google have implemented OpenID Connect, but find that users tend to just login using their Google identity. Also, OpenID Connect suffers from the same redirection problems, and typical OpenID Connect-enhanced providers still rely on username-password based authentication, and all personal data transfers are server-to-server. Lastly, OpenID Connect is currently managed by the OpenID Foundation, closely related to the Open Identity Exchange and NSTIC, and have thus not been through a proper global standardization process.

8.6 User-Managed Access

Attempts have been made to add additional privacy restrictions to OAuth via the User Managed Access (UMA) specification, which adds user-centric privacy capabilities to OAuth 2.0 (Hardjono, 2014). In particular, one critique of OpenID Connect and its backers is that they essentially get rid of the flexible metadata and policy approach of SAML while replacing it with an OAuth-flow that is optimized for large identity providers rather than user control, as users do not often understand how much personal data they are releasing via consent. In response, UMA creates an authorization interface for users to be aware of what OAuth data transfers are happening and what data a user controls. This work was started at the Kantara Initiative and is currently under exploration as part of NSTIC, and hopes to at some point be submitted to the IETF.

8.7 Messaging Standards

In general, one of the features of social networking is the movement of real-time data like status updates given as "activity streams" by sites such as Facebook and Twitter. While the traditional Web has focused on static content served via web-pages, the social web is moving towards a "real-time" web of heavily personalized content. While Twitter, Facebook, and Google all have proprietary methods for real-time updating in their social networks, there have been a number of proposed standards for enabling a real time web. The key of the real-time web in the context of decentralized social networking and social data stores is to dynamically update other nodes in the network based on social activities or the appearance of new data.

8.9 ActivityStreams

ActivityStreams is a serialization for the kinds of actions given in status updates in popular social networking sites. Social networking sites enable a number of actions, ranging from Facebook having users "like" items and Twitter sending "tweets" to followers. ActivityStreams works by a simple subjet-verb-object notation, similar to RDF, where there is an action by a user on another person or object, with an additional contextual target possibly involved. Each activity in a stream is given an explicit date in time, so an ActivityStream can be considered a stream of events in time, such as given by RSS and Atom feeds. (Atkins et al., 2010) While originally an Atom feed, ActivityStreams is now also serialized in JSON. A new proposed version of ActivityStreams, ActivityStreams 2.0, adds support for serialization using JSON-LD with arbitrary URIs for identifiers, and so becomes a more generic format for data with arbitrary schemas (Snell, 2014). ActivityStreams has widespread deployment, including the BBC, Facebook and Google, but is still new to many developers.

8.10 XMPP and Wave

The real-time Web for social networking is currently accomplished through two distinct architectures. The first architecture is founded on the IETF RFC XMPP (Extensible Messaging and Presence Protocol), which provides an XML "envelope" for data such as chat messages to be sent in real-time (St. Andre, 2010). Originally called "Jabber" and pioneered with the use of instant messaging, XMPP features its own standards for identity authentication and has an independent organization to host extensions such as those needed for decentralized social networking, XMPP Foundation. The XMPP Dialback proposal lets XMPP-enabled servers authenticate to each other in a decentralized way by using a verified key.[xxxv] As a mature technology that was up until recently used by GoogleTalk, XMPP can clearly scale to large decentralized networks. However, XMPP is not part of the Web (i.e. not built with HTTP) and so XMPP creates a whole parallel level of complexity with attendant security. Being originally designed for chat, XMPP requires very close to real-time synchronous persistent connections that may not be suitable for all social networking applications.

An early presentation by Evan Henshaw-Plath and Kellan Elliott-McCrea that hypothesized a decentralized Social Web suggested the use of XMPP.[xxxvi] Early federated social web application OneSocialWeb and BuddyCloud have been built on XMPPby attaching vCards and ActivityStreams to XMPP. [xxxvii]Perhaps the most advanced use of XMPP for decentralized social networking is Google Wave, which used the XMPP protocol for its identity layer and integrated chat using XMPP. Interestingly enough, Google Wave also innovated using XMPP to transfer updates of shared documents in near real-

time. However, Google abandoned Wave and has since focused on centralizing its identity eco-system around Google+. The Apache Wave project took on management of the Google Wave codebase, and a recent fork of the codebase by activists called Kune has begun work on the codebase again.[xxxviii] However, the fundamental protocols that Google Wave added to XMPP have not been through any standards body.

8.11 Pubsubhubbub and OStatus

In the audience for the original presentation on decentralizing the social web by Evan Henshaw-Plath and Kellan Elliott-McCrea was Roy Fielding, co-author with Tim Berners-Lee of the original HTTP protocol, who suggested that decentralized social networking would be better enabled by HTTP than XMPP. Brad Fitzpatrick, founder of the early social networking site LiveJournal and Google employee, went off to write up an HTTP alternative to XMPP for decentralized social networking that he called Pubsubhubbub (sometimes abbreviated PUSH) (Fitzpatrick et al., 2010). This HTTP-based architecture transformed the traditional "pull" HTTP architecture with a "push" architecture, similar to how RSS and Atom feeds worked. PUSH allows PUSH-enabled clients to authenticate to a PUSH server that they poll to receive notifications of activity. This allowed activity stream-based updates in near real-time over HTTP. In essence, when a node in a decentralized social network finds a friend they want status updates from, they subscribe to their feed via the "hub" of the federated social network. Then when the status is updated, subscribers are updated in near real-time when they poll the server. While this approach builds on the mature HTTP web and allows high latency and non-persistent connections, it is unclear if the polling architecture will scale. Also, due to is decentralized nature there is always the problem of "re-uniting" the conversation if a comment or response is made on another node of the decentralized social network. The Salmon Protocol attempts to tackle this problem by sending signed responses back to the original poster "upstream," so that the "upstream" node in the social network can re-create the conversation while keeping the entire framework decentralized (Panzer, 2010).

By combining ActivityStreams, PortableContacts, and WebFinger with Pubsubhubbub, Evan Prodromou – the original author of the free software clone of Twitter identi.ca – put forward the OStatus architecture (Prodromou et al., 2010). This architecture, based around creating Twitter-like functionality in a decentralized social network, provided the first documented meta-architecture for what Prodromou termed "The Federated Social Web." Prodromou produced an open-source version called Status.Net, and soon numerous other software coders were testing for interoperability using Ostatus's SWAT0 test-cases, including some commercial providers such as SuperFeedr. However, programming more complex features such as the Salmon Protocol and access control proved to be too difficult, and Prodromou went off to try to simplify the design with a new pump.io codebase. In the meantime, OStatus and Pubsubhubbub are maintained by W3C Community Groups.

8.12 IndieWeb

In the spirit of dramatic simplification, Tantek Celik – currently of Mozilla – coined the term IndieWeb to describe people using their own "independent" websites with domain names in order to control their own social data.[xxxix] Rather than attempting to create alternative decentralized social networks, the IndieWeb movement used as its motto to "create and publish content on your own site, and only optionally syndicate to third-party silos" such as Twitter and Facebook. Rather than ignore or simply take input in from silos, IndieWeb sites "push" data into them in order to reach the masses of users in

these silos. Also, the IndieWeb folks started focusing on areas previously ignored by the Federated Social Web, such as user experience and design.

Also, the Pubsubhubbub protocol was found to be too unwieldy, and so the IndieWeb advocates started working on an HTTP-based alternative called WebMention that got rid of the publishing hubs and simply recommends certain Link Headers on standard HTTP POST requests to let nodes in the decentralized social network send status updates to each other (Parcecki and Walters, 2014). This simple protocol should allow IndieWeb clients to communicate without creating much in the way of custom software and does not mandate the ActivityStreams format.

Interestingly enough, decentralized architectures should be able to use a vast variety of data-loads that go beyond status updates and include arbitrary dataloads in XML, RDF, or other formats produced even by sensor data. One new architecture that is using WebIDs in conjunction with a WebMention-style HTTP architecture for decentralized social networking is CrossCloud,[xl] a project by Tim Berners-Lee and Sandro Hawke that focuses on sending Linked Data around in a decentralized social network and possibly leveraging the work on the Semantic Web. However, this project has just started and so the ability to leverage data-stores in decentralized social networking that go beyond standardized personal data in profiles and activity streams is just beginning to be explored.

8.13 OpenSocial

While much of the work on decentralized social networking has focused on server-side protocols and data-formats, ultimately in order to be integrated into browser-based applications, an API to access this kind of data has to be standardized. Work such as ActivityStreams and PortableContacts drafted APIs for developers, but by far the most ambitious attempt to create a unified API for decentralized social networking was the OpenSocial API.[xli] A collection of Javascript APIs allowed Google Gadgets (a proprietary format for applications created by Google and also used by Google Wave, now abandoned by Google) to access ActivityStreams and PortableContacts as well as vCard for profile data. Although Google left the project, the open source Shindig implementation continues. As the Web moved to HTML5 and OpenSocial maintained a Gadgets and XML-centric architecture, OpenSocial is now incompatible as-is with HTML5, although it has found widespread success in creating intra-enterprise social networks in software produced by companies like IBM, SAP, and Jive where OpenSocial browser plug-ins can be mandated. As the OpenSocial specification was previously run by the independent OpenSocial Foundation, in response the OpenSocial Foundation has joined W3C and hopes to create a new version that drops its earlier XML and Google dependencies as well as being compatible with HTML5.

9. Socio-economic implications of social networking and data-driven identity ecosystems

While so far we have described the technical standards either existing or emerging in identity eco-systems, we need to understand the current socio-economic drivers of the growth of these eco-systems in context of the "big data" revolution in the market, and then re-frame the possible uses of these technologies in terms of social innovation. The evolution of the Social Web described in this deliverable has fostered in the last years the growth of innovative technologies and applications that emphasize mass scale user creation of content and wide participation to *harness users' collective intelligence*. Thanks to the proliferation of social media and smart devices, business innovation through digital platforms focuses mainly on the development of data-driven services, web and mobile applications based on "big data". The biggest Web players such as Google and Facebook, and online retailers such as eBay and Amazon, are contributing to the creation of the emergent big data industry (Schonberger and Cukier 2013). Big data is a way of describing the technical ability to collect and analyze in real time large data sets coming from people, sensors, and the environment that are stored in large data centers known as "cloud computing." Vast and growing amounts of data sets can be aggregated, stored, searched and correlated, containing organizational processes, personal information and personal location data (together with their metadata about the underlying information produced). This led to the development of cloud computing and sophisticated algorithms able to process this information and to discover new relationships among large data sets. Big data analytics is perceived to become a competitive advantage for innovative digital companies, becoming a key component in companies' digital strategy. Just to give an idea of the volume of data generated, 2.5 billion gigabytes of data are created every day (with Twitter alone generating t*erabytes* of data daily), and the data generated globally is expected to grow to reach 35 z*etabytes* in 2020 (Gantz and Reinsel 2012).

The core business model of the most competitive companies in these identity ecosystems is based on the firm's ability to extract value from social data (i.e. data produced and shared by users), and users generated contents through *mashup* processes, machine learning algorithms, and predictive analytics. Organizations are developing capabilities to derive insights from this information, automating decisions for real-time processing, identifying current and new business opportunities, identifying and predicting change, and quantifying current and potential risks. Access to vast data from heterogeneous sources, together with the computing power to process big data and the algorithms, is having a clear impact on many domains including medical science, logistics, healthcare, economic forecasting, retail, manufacturing, public sector and so on. For example, healthcare is already being transformed by far more feedback, including everything from real-time feedback on blood pressure to data on mortality rates and qualitative reports of patient experience. It is clear that organizations are increasingly becoming more dependent on big data development for critical decisions and applications (Kuner et al. 2012). Knowledge of machine learning are applied in commercial environments by making sense of large pools of data, such as employing pattern recognition technologies that look for hidden patterns or anomalies in large datasets, detecting data clusters and discovering new correlations, predictive patterns,

and real-time statistical modeling. Examples include designing predictive systems based on mass user behavioral analysis and using *social sorting* techniques that assign users to a particular category. This can enable the system to make predictions about future user behaviors based on the past user experiences, and therefore adjusting their future experiences accordingly with the aim of maximizing sales and profits.

Companies are creating goal-oriented and personalized applications that *mashup* personal information, location data (GPS, micro-geographical), social graph (likes, friends, locations, posts), behavioral analytics (movement and duration on the webpage), people's e-commerce shopping behavior, device data (apps installed, operating systems) and data and data coming from sensors in order to provide customized services to customers. In this way companies are able to identify innovative business and revenue models to capture value and increase profits from the collective intelligence of the users. This trend will only increase, as technologies such as the RDF-based Linked Data stack are not heavily used yet, but aim to increase the kinds of mash-ups capable of being done by companies by using URIs as universal identifies. The collection of personal data happens easily with the explosion of mobile connectivity, since companies are able to aggregate data about browsing behavior through the use of *cookie*s, or using the *Facebook Like button* and create individual and collective profiles that cluster all these activities (Roosendaal 2011). Furthermore, many new applications base their business model on being able to aggregate personal data and localize the physical position of users collecting real-time GPS location and DNS lookups. Many Internet users are unaware that data are being collected, analyzed, mined and sold to advertisers (McDonald and Cranor 2010).

For instance, Facebook tracks and traces users and processes their data, since 23.1% of all online advertising not on search engines, video, or e-mail, run instead on Facebook. But at the same time, data-mining companies and data aggregators are "scraping" personal data that is publicly accessible to sell them to third parties, such as data aggregators and credit agencies (Van Eijk et al. 2011). Moreover, *Facebook Connect* is able to keep tracking users across the Web, bringing rich data on consumers' behaviors back to the Facebook platform where they can be controlled and monetized in the emerging personal data marketplace. Behavioral advertising "entails the tracking of online behavior of Internet users in order to build a profile of these users to target them with customized advertising" (Van der Sloot and Borgesius 2012, p. 2).

Google competes with Facebook, making its revenue mostly from advertising, and after acquiring Double Click in 2007, has been focusing on behavior advertising through its "Interest Based Advertising" program. Google thus tracks the behaviors of its users through cookies across its advertising network. Google can enrich users' behavioral profiles with additional data gathered from other Google services and with information that users upload to other social networks. Within this type of personalised business model innovation *lead users* are also recognised as sources of innovation in new product development (von Hippel 1986, 2005). The role of user communities as co-producers, including open-source development and online communities constitute the foundation of social media platforms that companies are increasingly able to capture to appropriate economic value (Franke and Shah 2003; Thomke and von Hippel 2002).

This process of big-data analysis and machine learning that tightly integrates innovation in business and technology is now mastered by a very small group of US-based companies, which in turn have re-centralized the Internet and of the intangible value creation process by fostering the rise of natural monopolies and dominant positions – mostly due to seizing control of personal data and identity eco-systems. Today the Internet as a whole, is becoming increasingly concentrated with Google controlling

nearly 82% of the global search market and 98% of the mobile search market, Facebook dominating the social networking and identity ecosystem, while Apple, Amazon and Microsoft controlling the mobile market and cloud-based services platforms. The global telecom operators are also expanding their business towards this new market, attempting to regain control on the cloud services that are conquering the majority of the market shares, and proposing to tax "over the top" players, representing a clear threat to net neutrality. Such players are able to seize network externalities resulting from economies of scale and network effects that generate increasing returns associated with their lock-in strategies. As shown by relevant economic and innovation research on multi-sided platforms (Gawer and Cusumano 2002; Gawer 2011; Gawer and Cusumano 2012) social networking platforms, such as Facebook are platform owners that value user mobilisation as a priority to maximise their profits based on the exploitation of the network effect created by social graph and personal data of its the huge user base. Thus, Facebook encourages a critical mass of adoption, while monetising the installed base through advertising (Boudreau and Hagiu 2009). Then the secret for profitability and growth is to activate the social graph, by keeping linkages among active members and facilitating and imposing engagement and interaction on the platform. In addition, Facebook has designed a marketplace for ecosystem innovation based on applications built by a community of 600,000 developers. Many applications and widgets built on the Facebook platform are inherently social in nature, and they lead to building, activating and refreshing the social graph by enhancing network effects and attracting new members to the platform (Boudreau and Hagiu 2009).

10. Who owns the Data in a Data-driven society?

The NSA "data-gate" showed that governments are engaging in vast surveillance operations and the US National Security Agency is acting as the ultimate "big brother" personal data aggregator. Classified documents leaked by the former US security contractor Edward Snowden confirmed that the NSA does not need any court authorization to secretly collect and sift through all contents of Internet communication, phone calls and store large amount of personal data about citizens with a worldwide scope, with huge implication on civil liberties, data protection, and privacy (Bria et al. 2013). This implies that Google, Facebook, Yahoo, Microsoft, Skype, and other Silicon Valley firms and major phone networks are not only centralizing the Internet, but providing open-ended access to users' data, which provides a technical foundation for US authorities the ability to spy on people's private data. Clearly, these revelations are creating a debate amongst the public, eroding public trust in Internet governance, and raising concerns that the open and transparent Internet of today risks to become a market of citizens' data, dominated and managed by few monopolies that facilitate the overall surveillance of the global Internet by intelligence agencies.

The latest commercial trends towards a concentration of actors, vendors and non-standardized data lock-in is preventing a standardized, privacy-aware, and open identity eco-system. The current identity eco-system is still a "Wild West" in legal terms, as companies even sometimes illegally selling users personal data to third parties, or personalizing services by using personal information without explicit user's consent, always putting the onus on users to opt-out, as opposed to asking them if they want to opt-in. However, without a proper legal and technical framework, the trend towards new business models in the Internet ecosystems based on "personalization" are supported by the advertising industry, incorporating the private life of users in the marketing process, will continue unabated: Personal social data is the new profitable market. Users' "social graphs" (friends and relationships) and "interest graphs" (what people like and do) are harnessed and sold to advertisers to extract and 'mine' targeted market information. This model exploits users' personal information to deliver targeted advertising, service and social recommendations through collective filtering and semantic data analysis, resulting in a "filter bubble". This trend is very clear in location-based services and in life style apps where the geospatial information of users and sensitive information about users' and their social networks' tastes, relationship and interests are analyzed and aggregated to create personalized offers. However, despite the large profit margins and dubious legality of many of these business practices, users receive only free or cheap services in return for their data.

In fact, users of these ecosystems don't own their personal and social data, but they end up renting their own data from Facebook and other free social networking services after they gave these companies all their rights by signing their legal terms of services. One of the issues is that users usually often can't understand their consent to terms-of-service legal contracts, and these contracts often cannot be enforced technically. They then lose control of not only explicit personal data, but implicit data in the digital trails that users leave around the Internet with their searches, purchases, uploaded content, and conversations. This data exhaust is the personal data that companies collect about what products their customers buy and how they use digital services. In this way businesses are acting as brokers of personal

and sensitive data that are manipulated through subtle privacy infringements mechanisms. For instance, with the implementation of Facebook's "frictionless sharing" policy and the various app services that rely on centralized commercial clouds, people are showing growing concerns about these commercial practices, urging authorities to update privacy and trust regulations. Locking in users' social data is creating a new "data enclosure" that consists on capturing users' co-created value through network or device lock in, segmenting the network in other areas and overruling the network regulations by imposing their governance models.

As a consequence, the most important data that is becoming available on a vast new scale is information about people's behavior. For instance location data from mobile phones and evermore consumption data as people increasingly use credit cards for purchases. This fine-grained behavioral data, change the way we think about society and even how a society is governed, since we can track social phenomena down to the individual level, and the socio-economic connections among individuals. Decisions made about the types of data provided, the degree to which users can modify such data, and the scope of access they are afforded will have an enormous impact on the ways in which people access services and knowledge, and in the way authorities justify their decisions based on the analysis of big data in real time. The situation as regards user-generated knowledge, data and users' identity on the Internet, is largely unregulated, resulting in an "identity marketplace" that risks to generate a huge "privacy bubble". Personal and social data are the new currency, as shown by the recent Facebook and Twitter IPOs and the recent reports and discussion at the World Economic Forum that are grounded in the conception of personal data as a new asset class, an even larger Internet-based personal data market is currently in the making. Furthermore, with the rise of the so-called Internet of Things (IoT), also objects, machines, and non-human organisms actively provide information into this identity ecosystem.

11. Ethical implications of current social networks on Privacy and Data Protection

Social networking presents novel opportunities together with big challenges especially to privacy, identity management, and data protection. Researchers emphasize the risk that Web 2.0 is posing to privacy and fundamental freedoms, by defining it as the next privacy bubble (Scott 2008), and advocating the need for new privacy enhancing technologies that give users control over their personal information (Langheinrich 2001; Stalder 2002; Cavoukian 2009; Hildebrandt and de Vries 2013; Solove 2006; Wright and de Hert 2012). Open-source software communities have been very critical of some of the latest development of the Web 2.0. In particular, Tim Berners-Lee (2010, 2012) has published a series of articles alerting to the danger of the Web "walled gardens", by referring to centralized platforms, such as Facebook and Twitter, that don't allow open standards and data portability. Creating "walled gardens" out of personal data means that when users try to leave the platform or to move their data to another service, they lose everything, including the social relations, profile data, and the possibility of communicating with their friends. According to privacy researchers, too little space has been given to the critical exploration of the impact of Web 2.0 on society, ethics, and the economy. A particular emphasis is on the threats that social media poses to the concept of privacy and individual identity, and how sensitive personal data flow across networks are being extracted for commercial gains (Burdon 2010; Dwyer et al. 2007).

Surveillance researchers are currently looking at the implication of social media and digital technology on personal freedom and identity (Lyon 2002, 2007; Rodotà 2012). Surveillance technologies are currently very pervasive. Every day we encounter CCTV cameras in the street, we use rfid-enabled credit cards, navigation systems, machine readable passports, we constantly use mobile devises and surf the Internet for an increasing number of activities, we have a health card number, loyalty cards and so on. Personal information is gathered by many means (e.g. digital, biometric, genomic), and are constantly recorded, stored, retrieved, compared, mined, traded and processed. As explained before, today most of the social applications are created by aggregating personal data. An increasing number of employers use social media in addition to the traditional résumé. Some employers are even asking applicants if they can have their access to their Facebook passwords. In this way social media companies, together with data brokers and aggregators are creating new categories, and personal profiles, so they can build a composite picture of their users, asking them to contribute directly to those categories and to the categorization, aggregating their preferences, musical tastes, food preferences, political affiliation, religious commitment and so on. Surely those categories don't necessary fit the way people want to present themselves and through these categories risks and opportunities are assessed, and people's life-chances and choices are influenced and managed in a process that has been named "social sorting" (Lyon 2002).

This process of data aggregation and the data marketplace that exchange users' personal information often happens without explicit users' consent, thus posing a serious threat to privacy and data protection online (Bauman and Lyon 2012; Ball et al. 2012). Various solutions have been advocated by academics and advocacy groups, such as the need to follow a different technological approach based on open standards, including data portability, privacy, and user control of personal data. Zimmer argues that

building personal profiles of large numbers of people to predict their behaviors and deliver targeted advertising is creating an unfair bargain, resulting in an infrastructure of control more than an infrastructure of empowerment (Zimmer 2008). The amounts of raw social data gathered through sophisticated consumer tracking algorithms allow for detailed analysis of users' behavioral patterns. For instance, according to Ibrahim (2008), Facebook acted as social norm setter, by transforming the very meaning of privacy through enabling new users' behaviors around "frictionless sharing" on its platform, without their explicit consent. The very concept of privacy is transformed, because the more people share data about their personal life, the more they are rewarded with peer attention, which is a strong social motivation to continue to share information and aggregate trust and reputation. This is coupled with platform architectures designed to lower privacy levels and influence users' perception of risk (Ibrahim 2008).

However, these kind of positions that victimize the users and condemn the company have been criticized. For instance, it has been argued that Facebook seduces people with social motives into using its services; "Facebook succeeds in feeding an individual's longing for self-promotion and identity management, seducing them into accepting the trade-off between their privacy and perceived benefits of social capital" (Oosthuyzen 2012). Certainly the new norms and behaviors of social media providers are challenging existing legislations, especially in Europe. In the European Union there are several kinds of legislation such as the Data Protection Act that protect privacy regarding the processing of personal data. This legislations are being currently rewritten to answer the new challenges posed by digital communication and social media (Van der Sloot and Borgesius 2012; Wong 2013).

Other critiques of the current social networking paradigm stem from a more philosophical perspective, outlining the limitations and rigidity of the Facebook approach in defining identity and social relationships. Researchers looking at social and psychological effects of social media are analyzing the way social media mediate social relationships and identity, talking about a possible collective "disindividuation effect" (Stiegler 2010). The individual in fact cannot be conceived outside the holistic relation between the individual, the collective, and the technological systems that constitute the environment (Virno 2008). If identity formation within social networks remains solely based on engineering processes, it can create personality mismatch and artificial social interactions (Hui and Halpin 2012). Other dysfunctional effects created by current social networking are investigated within the emergent discipline of cyber psychology (Riva and Galimberti 2001; Riva 2005). This new discipline analyses processes of change activated by social media that are centered on social interaction. This approach shows that social media present at the same time new opportunities and new problems, creating new types of empowerment and collective activation, as well as new dysfunctional behaviors. For instance, psychologists are analysing the identity crisis due to multiple identities that young people construct when they are using the Web, and that can interfere in an integrated development of youth social identity (Riva 2010). Moreover, excessive use of social media can create "Internet addiction disorder" and "emotional illiteracy" (Barbera et al. 2009; Kuss and Griffiths 2011) which created a new wave of psychology to study these kinds of pathological behaviors named "New Media Related Psychopathology" (Morahan-Martin 2000)

12. Conclusion: A European alternative

A main Internet trend-threat that is recognized within the current situation is the increasing concentration of power in the hands of a few data aggregators, none of which is located in Europe. What is needed in Europe and is a new sort of socio-legal-technical framework that demonstrates that privacy-enhancing technologies are compatible with viable, sustainable and innovative business models, with effective security requirements and generally accepted performance standard. One key question is how to assure user control over personal information in an ocean of commercially valuable Big Data: Technical solutions do not work by themselves, therefore legal and commercial solutions have to be based in technology and integrated with the appropriate policy framework.

The overall aim of D-CENT is to research new digital structures and bottom-up regulatory mechanisms that can enable new types of consensual governance that are more flexible, and effective than existing structures and which evolve over time through user participation, much as the World Wide Web, open source software, and other network-based ecosystems have evolved in unexpected ways through bottom-up mass participation. The question that D-CENT will address is how to build future socio-economic models and governance frameworks that emulate open-source and Creative Commons licenses in enabling collaborative value to emerge more readily and in ways that are shareable and protectable. A multidisciplinary approach is needed to integrate technology and information policy within a broader understanding of normative, regulatory, institutional and governance aspects. D-CENT will carry out a broader investigation linking technologists with social scientists, citizens, activists, and policy makers to understand the implication of the current identity markets and mechanisms that are crucial for the understanding of the future immaterial economy and financial trends and for the design of alternative effective instruments of social interaction, democracy, and economic exchange (D 3.3).

An alternative framework is needed to provide an open architecture for managing online identity, security, data, and collective governance in an integrated fashion and based on democratic and participatory processes (Bria, F. 2013). The US Government has been proposing the new self-regulatory framework for an "identity ecosystem" in NSTIC to develop a new network architecture and market for personal data that will enable new forms of trusted governance, and secure commercial transactions. The process is driven by large US corporations such as Paypal, Google, Verizon, Citygroup with a focus on economic gains, thus betraying a multistakeholder approach where citizens and civic society are represented. The US framework is cannot be reproduced in Europe with different data protection, privacy and policy frameworks. The question is to create open public ecosystems rather than a winner-take-all marketplace whose dominant players set the terms of innovation and competition (Bria, F. 2012).

From a regulatory standpoint, momentum is slowly moving towards more coherent support for privacy in Europe, for instance with the Data Protection Regulation. Progress should be made on "data portability", proposing that people should have the right to obtain copies of their data – in an interoperable format – so that they can more easily change from one service to another. Furthermore, a key question is how to protect individuals from profiling of their personalities and behavior by companies, without their explicit consent, thus enforcing the control of the individual over their personal data, and of communities of their social data. In the European Parliament there are discussions

over profiling techniques such in the EU Cookie Directive and the W3C "Do Not Track" (Tracking Protection) standard for privacy online. Questions around data protection, online identity and privacy will become increasingly important in the Internet of Things that has already been defined as a post-privacy situation in which people, computers, the Internet and objects jointly form connected networks with full traceability and within a continuous stream of data.

D-CENT aims at identifying shared practical solution by developing alternative distributed and decentralized technology tools, based on open standards for instance to social networking, that should be extended also to data searches and data storages (distributed clouds). However, it is difficult for solutions to emerge in approaches that do not reserve a clear advantage for the developers, which instead is the case for centralized solutions. In a decentralized social network there is no single central entity that access and control the data of all subscribers, as it is the case in Facebook. A strong public intervention at EU level could support this alternative area of development that is far from being within the short-term interests of even large EU industries, and so far has been left to isolated developments, activists and hackers, and users themselves. Recognizing its strong social value, an alternative identity eco-system based on decentralized social networking would allow a whole new generation of industrial and social innovation to start in Europe. There is a clear emerging need of supporting the development of a distributed architecture as a kind of "regulated monopoly" able to ensure basic digital services at European level, on top of which a whole new open ecosystem of services and applications could flourish, in a participatory bottom up and user-driven innovation model, based on open source and open hardware developments.

Furthermore, the dominant business models in the market today don't seem to be sustainable for the European economic system that consists of 99.6% of SMEs and has a very different equity and venture capital market compared to the US. There are major differences between European and US economic structures, especially regarding the absence of big firms that innovate on the data layer and the structure of venture capital network to fund technology start-ups. In Europe there are other actors that should be supported in order to drive innovation. In the first place, infrastructure should be built together with users and civic society actors, cities and regions that are closer to citizens and SMEs and they can more easily engage them into the innovation process, applying methodologies, incentives and policies to facilitate their involvement, making sure that services deployed answer to concrete local needs and demand. The governance structure for this alternative models could be a mixed model based on local communities or other civil society actors, acting as intermediaries (for interesting European examples grounded on open and interoperable technology solutions where local governments have a key role in the governance structure see projects as Commons4Europe[xlii] or CitySDK[xliii] funded by the EC within the CIP framework.

A classical P2P fully distributed solution relies on personal devices to store the data, and on individuals to correctly manage them. Which is the reason why fully centralized solutions, such as Google or Facebook, are so successful: the user has not to worry about storage of data, or its treatment. It is however increasingly difficult to create trust between users and institutions, when institutions are very opaque about what they actually do with users 'data as shown in the Prism case. The decentralized and federated solution that can be envisaged in the context of D-CENT is instead one where the citizen entrusts a local authority, institution or community organization (such as a city council, or a university, a citizen foundation, or another third party not necessarily "local" in geographical terms) to manage their data, not only as for its preservation and accessibility in a local cloud (under local jurisdiction control),

but also for negotiating the usage of such data for any other commercial purpose, maximizing the value citizens can gain beyond monetary incentives. In order to emphasize the benefit of these alternative non-market models it is necessary to move from a transactional paradigm that sees personal data as a new "asset class" to a relational and ecological paradigm that consider social data as commons that can valorize the social cooperation of communities and re-appropriate the collective value generated in diverse socio-economic and cultural contexts. This requires transforming personal data to social data with the appropriate open technical standards for access control.

12.1 Gap Analysis of Existing Standards

In order for a new European decentralized social networking eco-system to commence, there must be a solid basis in open standards. However, the current fragmented standardization landscape lends itself to support the use of proprietary technologies by large vendors and so excludes SMEs and open-source projects who may have difficulty with licensing or implementing the protocols necessary. Luckily, the W3C is stepping in and is currently in the process of creating a new Social Web Working Group to standardize key architectural components of a decentralized and federated social web such as ActivityStream and OpenSocial, an effort that is to be chaired by IndieWeb founder Tantek Celik of Mozilla.[xliv] This should provide the licensing and patent commitments needed by open-source projects like D-CENT, and it is expected that via the W3C and Neo, D-CENT will actively participate in the Social Web Working Group.

Still, a number of outstanding problems have been revealed by this analysis of the social standards landscape. First, there is no clear user-owned identification scheme, and thus by default user's identifiers such as URIs and emails are usually under the control of private companies. While a long-term solution in this is an area for future legal and socio-economic research on naming and the IANA function, it can be solved in the short-term in the D-CENT project by having citizens groups and public bodies purchase long-lived domain names that they can then legally protect for user's identity. Then the actual personal data of users themselves can be stored using extensible modifications to commonly used standards, with a focus on the widely deployed vCard, which as a baseline standard for data portability should allow the users to move their profile and personal data from one D-CENT platform to another, so not "locking" user's personal data into one particular identity provider. The D-CENT project should encourage the use of upcoming W3C standards such as ActivityStreams and OpenSocial This should provide the technical foundation for a much higher-degree of user choice and trust in citizen-centric identity eco-systems than currently allowed in existing corporate identity eco-systems.

Another gap is security and user consent of data transfer. While this topic is also an outstanding issue for researchin decentralized social networking systems, a number of short-term standardized solutions do exist. Currently, the landscape of authentication technologies is also biased towards either low-security username-password combinations versus high-security enterprise environments. Attempts such as WebID and BrowserID to bring higher-security authentication to ordinary users have to a large extent failed. Authorization technologies such as OAuth work today, but need a higher degree of user-control. However, with the advent of the Web Cryptography API that unifies cryptographic functions in Javascript across browsers in combination with legislation such as the eSignature directive that revises liabilities and notifications for breaches may present sufficient initiative for having ordinary user's engage in better authentication practices, while the work on the UMA presents an attempt to use OAuth technologies in a decentralized environment while preserving user autonomy and privacy. The simplest

way forward in the D-CENT project is to encourage the use of digital signatures in any authentication and authorization process in addition to user-name passwords. This can likely be accomplished by having users in the D-CENT project attach public keys to their user-names using simple technology such as WebFinger, and then use these keys to sign authentication and authorization requests. This still leaves the problem of secure private key transfer between multiple devices, but future work in the W3C Web Cryptography API and open-source projects such as LEAP[xlv] may provide valuable design patterns to be used by the D-CENT projects. Lastly, the maturation of OpenSocial and ActivityStreams inside of the W3C should allow users to embed identity eco-systems of their choice within HTML5-based applications, which gives users the ability to personalize their interactions with the Web without having to use one of the large identity silos.

The overall architecture of the D-CENT decentralized social network should aim at a move from personal data to social data. However, current architectures in both the HTTP and XMPP realm are based on an emphasis towards individual data with all data being public by default, not the group-based data control as likely needed by citizen groups and public institutions that respect privacy and may have their own group-based social contracts with their users. Currently, none of the architectures except very basic and non-standardized work around access-control lists with Google Wave tackle this problem of defining "groups" in a decentralized social web. Rather than simply regard every decentralized social network as a network of users with their own personal data, it would be far better to regard communities as "first-class" citizens in the social network, with their own activity stream and social data stores. This can already be seen in the fact that not only individuals have Twitter accounts, but often groups such as Anonymous or companies such as BP have their own official Twitter accounts. This would require not only access control rules or capabilities based on community-based social contracts, but also some fundamental re-thinking of the architectural assumptions of both IndieWeb-based and XMPP-based approaches, as individual personal data formats such as vCard would have to be replaced with much more open social data formats. These social data formats should allow generic kinds of data, as each kind of community will have often unique and unpredictable needs for data that go beyond the kinds of formats optimized for Facebook profiles. Luckily, ActivityStreams 2.0 is moving to a more generic based open data environment that should allow arbitrary data payloads to be carried by streams and then integrated into social data stores.

Thus, the D-CENT project should investigate supporting generic social data stores in a type similar, but less individualistic, than the CrossCloud project. D-CENT should support developer-friendly JSON, and when possible Linked Data-style formats based on JSON that offer the kinds of extensibility necessary for meeting very particular community requirements. Overall, the thesis that moving from individualistic personal data stores to social data will increase the value of the data requires new work in technical standardization in the Social Web Working Group, but may have huge implications. Taking the example of Metcalfe's law that states that the value of a network such as a social graph grows proportionally to the square of the number of connected users of the system (n2), it has been hypothesized by Reed that in group-forming networks the number of possible groups increases exponentially proportional to 2n.[xlvi]This means that these kinds of communities should be able to produce even more value than individuals in a social network.

Interestingly, one of the areas left out of the current standardization landscape is digital currencies. The original work on digital currencies by Chaum as eCash was riddled with patents and licensing issues and so never took off (Chaum, 1985). However, with the advent of popular cryptocurrencies such as

Bitcoin[xlvii] – which keeps track of a ledger of all Bitcoin mining (the process of converting energy into Bitcoin via repeated application of cryptographic operations) – there is renewed interest in digital currencies from all sectors of society. Furthermore, alternative Bitcoin-like frameworks have developed as developers create their own append-only ledgers for new kinds of currencies, such as the meme-turned currency Dogecoin.[xlviii] There is even interest in transferring traditional fiat currencies using similar techniques via open-source frameworks like Ripple.[xlix] While there are no IETF or W3C open standards in this space, the W3C is hosting a workshop on digital payments[l] and the D-CENT project is expected to participate and engage with whatever open standards result from the work in its effort to create a sustainable community-based digital currency for the social good.

12.2 Towards a new Ecological and Ethical paradigm for Data access and exchange beyond transaction and monetization

The new ecological and relational paradigm of community-based social data presupposes a strong focus on ethics and collective awareness. For instance DCENT could develop and test together with users a *European ethical digital rights guarantee* that could be applied to the way data, identity, security is managed within digital ecosystems. In Europe we need an alternative framework for data ownership, use, and exchange that is embedded within the context of European values, ethical guarantees and regulations. D-CENT aims at developing a framework that demonstrates that privacy-enhancing technologies are compatible with viable, sustainable and innovative business models, with effective security requirements and generally accepted performance standards. A focus on ethics can also be connected to a bigger effort in promoting useful evidence in decision making, thus highlighting the challenges faced in embedding evidence into the decision making process.

The new ecological and ethical data paradigm advocated here should be framed in the context of a wider economic movement that is trying to reformulate the narrow economic indicators that currently measure and evaluate performance and wellbeing. Conventionally, economic impact is measured in terms of income or employment. However, economic multiplier effects for the social economy, including digital and complementary currencies are increasingly widespread with the idea that, beyond economic and financial indicators that measure economic development, productivity and output, (such as Gross Domestic Product per capita) there is a need for novel indicators that measure intangible value and wellbeing. Over the years, at the light of the financial crisis and of the inability of mainstream economic approaches to value intangible assets, new indicators and indexes have been compiled by researchers, NGOs and governments. These new indicators emphasize the importance of going 'beyond GDP' in order to find new ways to measure sustainability and a comprehensive set of values.

For instance, the Stiglitz-Sen Commission was set up to investigate beyond GDP indicators and amongst other issues recommended that (1) When evaluating material well-being, look at income and consumption rather than production (2) Consider income and consumption jointly with wealth (3) Give more prominence to the distribution of income, consumption and wealth (4) Broaden income measures to non-market activities (5) Steps should be taken to improve measures of people's health, education, personal activities and environmental conditions. Following the trends established by the Stiglitz-Sen commission, in May 2011, on the occasion of its 50th anniversary, the OECD launched an interactive indicator "for a better life", intended to measure the well-being of people and better understand their

living conditions. Another example is the Happy Planet Index launched by the UK based New Economics Foundation in 2006 that has been carried on by Nesta with the launch of Action for Happiness,[li] a movement for positive social change that promote transformative, participative and action oriented activities to promote better social connections and social support. The examples above show the shift globally in moving from measuring material growth to measuring the things that matter most, adopting a holistic approach that takes into account community wellbeing, labor conditions, environmental sustainability, and democratic participation.

Exploring new metrics for measuring the non-financial value of social innovation and immaterial goods, can be also applied to the social value of data and intangible resources that communities exchange within Digital Ecosystems. D-CENT will build on practical experiences, and looking on how to apply alternative valorization models in practice, through innovative mechanisms for sharing intangible common resources based on community ownership of social data, trust mechanisms, and crypto digital social currencies. This is in line with the overarching objective of D-CENT to promote new forms of digital social collaboration, sustainable consumption and exchange, by enhancing cultural practices embedded in processes of production, consumption, and exchange of goods and services based on community collaboration. D-CENT will therefore be grounded on the "commons" model, as a governance structure to negotiate rules and boundaries for managing the collective production and access to, shared resources. Governing of the commons honors participation, democracy, transparency, equal access, and long-term sustainability.

The perspective of the "commons" will be adopted to understand decentralized digital ecosystems where intangible resources are shared pubic goods and their exchange aims at maximizing social impact and citizens' engagement. The core of the work D-CENT will do in this area will be linked to the implementation of the economic analysis of WP3 (D 3.2; D3.3; D3.4; D3.5) and formulated within a socio-economic and impact assessment framework in WP1 (D 1.3). The work of WP3 in synergy with WP1 and WP4 will then provide the socio-economic framework and analysis that will be used as a basis on which to construct novel indicators. Concepts such as self-governance of the commons and alternative economic cultures will inform the analysis and will underpin a novel concept of social data as commons promoting a high level of community ownership, privacy, access and openness by design.

---

i https://www.youtube.com/watch?v=QOEMv0S8AcA

ii http://open-stand.org/

iii http://www.ietf.org/about/note-well.html

iv https://www.rfc-editor.org/rfc/rfc5378.txt

v https://www.rfc-editor.org/rfc/rfc4879.txt

vi http://www.w3.org/Consortium/Patent-Policy-20040205/

vii http://www.w3.org/2005/10/Process-20051014/

viii http://twister.net.co/

ix http://about.psyc.eu/Protocol

x https://www.w3.org/2014/strint/

xi https://www.youtube.com/watch?v=oV71hhEpQ20

xii https://gnunet.org/

xiii http://secushare.org/

xiv http://leap.se

xv http://www.spion.me/

xvi https://abc4trust.eu/ and http://primelife.ercim.eu/

xvii https://research.microsoft.com/en-us/projects/u-prove/

xviii http://www.zurich.ibm.com/security/idemix/

xx http://danbri.org/words/2008/01/29/266

xxi http://pde.cc/

xxii http://www.nist.gov/nstic/

xxiii https://www.eid-stork2.eu/

xxiv http://www.icann.org/

xxv https://www.iana.org/

xxvi http://1net.org

xxvii http://namecoin.info/

xxviii http://telehash.org/

xxix http://openidentityexchange.org/

xxx http://kantarainitiative.org

xxxi http://codebutler.github.io/firesheep/

xxxii http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security

xxxiii https://github.com/mozilla/id-specs/blob/prod/browserid/index.md

xxxiv http://openid.net/connect/

xxxv http://xmpp.org/extensions/xep-0220.html

xxxvi http://www.slideshare.net/kellan/beyond-rest

xxxvii http://buddycloud.com/

xxxviii http://kune.ourproject.org/

xxxix http://indiewebcamp.com/

xl http://crosscloud.org/

xli https://github.com/OpenSocial/spec

xlii http://commonsforeurope.net/

xliii www.*citysdk.eu/*

xliv http://www.w3.org/2013/socialweb/social-wg-charter.html

xlv http://leap.se

xlvi https://www.princeton.edu/~achaney/tmve/wiki100k/docs/Reed_s_law.html

xlvii https://bitcoin.org/

xlviii http://dogecoin.com/

xlix https://ripple.com/

l http://www.w3.org/2013/10/payments/

li http://www.actionforhappiness.org/

Bibliography

Atkins, M., W. Norris, W., Messina, C., Wilkinson, M. and Dolin, R. 2010. Activity Streams Concepts and Representations,. http://activitystrea.ms/head/json-activity.html

Ball, K., Lyon, D., & Haggerty, K. (Eds.). 2012. Routledge handbook of surveillance studies. Routledge.

Baran, P. 1962. On Distributed Communications Networks. RAND Corporation papers, document P-2626.

Bauman, Z., & Lyon, D. 2013. Liquid Surveillance: A Conversation. Polity.

Recommendation, W3C. http://www.w3.org/TR/1998/REC-xml-19980210

Berners-Lee, T. 2006. Linked Data. http://www.w3.org/DesignIssues/LinkedData.html

Blair, A. M. 2010. Too much to know: Managing scholarly information before the modern age. Yale University Press.

Boudreau, K. J., & Hagiu, A. 2009. Platform rules: Multi-sided platforms as regulators (pp. 163-191). Cheltenham, UK: Edward Elgar Publishing Limited.

Bray, T., Paoli, J., and Sperberg-McQueen, C. 1998. Extensible Markup Language (XML).

Bria, F. 2012 Open Innovation Service Policy Group (OISPG) Service Innovation Yearbook 2012

Bria, F. 2013 Kynote,iSamenleving symposium, Amsterdam 14th February

Bria et al. 2013, Internet Identity Ecosystem Seminar, Rome Conference position paper

Brickley, D. and Miller, L. 2014. FOAF Vocabulary Specification .99. http://xmlns.com/foaf/spec/

Burdon, M. 2010. Privacy invasive geo-mashups: privacy 2.0 and the limits of first generation information privacy laws. JL Tech. & Pol'y, 1.

Cavoukian, A. 2009. Privacy by design. Take the Challenge. Information and Privacy Commissioner of Ontario, Canada.

Chaum, D. 1985. "Security without identification: Transaction systems to make big brother obsolete". Communications of the ACM 28 (10): 1030.

Clark, D., 1992. A Cloudy Crystal Ball -- Visions of the Future. Presentation given at the 24th Internet Engineering Task Force. 1992-07-16.

Clayton, N. 2011. Whatever Happened to Diaspora the 'Facebook Killer'? Wall St. Journal. Nov 7, 2011. http://blogs.wsj.com/tech-europe/2011/11/07/whatever-happened-to-diaspora-the-facebook-killer/

Crockford, D. 2013. JSON. ECMA. http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf

Danezis, G., Díaz, C., Troncoso, C. and Laurie, B.. 2010. Drac: An Architecture for Anonymous Low-Volume Communications. Privacy Enhancing Technologies 2010: 202-219.

Dwyer, J. 2010. Four Nerds and a Cry to Arms against Facebook, New York Times, May 12th 2010. Available at http://www.nytimes.com/2010/05/12/nyregion/12about.html

Dwyer, C., Hiltz, S. R., & Passerini, K. 2007. Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace. In AMCIS (p. 339).

Fiege, U., Fiat, A., and Shamir, A.. 1987. Zero knowledge proofs of identity. In Proceedings of the ACM Symposium on Theory of Computing (STOC '87), Alfred V. Aho (Ed.). ACM, New York, NY, USA, pp. 210-217 (1987).

Fitzpatrick, B., Slatkin, B. and Atkins. M. 2010. Pubsubhubbub Core.

Franke, N., & Shah, S. 2003. How communities support innovative activities: an exploration of assistance and sharing among end-users. Research policy,32(1), 157-178.

Franks, J., Hallam-Baker, P., Hostetler, J., Lawrence, S., Leach,, P, Luotonen, A. and Stewart, L. 1999. HTTP Authentication: Basic and Digest Access Authentication. IETF RFC 2617.

Gantz, J. and Reinsel, D. 2012. The Digital Universe in 2020: Big Data, Bigger Digital Shadow s, and Biggest Growth in the Far East, IDC iView December 2012, sponsored by EMC

Gawer, A. (Ed.). 2011. Platforms, markets and innovation. Edward Elgar Publishing.

Gawer, A., & Cusumano, M. A. 2002. Platform leadership (pp. 252-254). Boston: Harvard Business School Press.

Gawer, A., & Cusumano, M. A. 2012. How companies become platform leaders. MIT/Sloan Management Review, 49.

Guha, S., Daswani, N., and Jain, R., An Experimental Study of the Skype Peer-to-Peer VoIP System, IPTPS 2006.

E. Hammer-Lahav and Cook, B. 2011. Web Host Metadata. IETF RFC 6415, June 15th 2010. http://tools.ietf.org/html/rfc6415

Hammer-Lahav, E. 2010. The OAuth 1.0 Protocol. IETF RFC 5849. http://tools.ietf.org/html/rfc5849

Hardjono, T. 2014. User-Managed Access (UMA) Profile of OAuth 2.0. IETF Internet Draft.

Hardt, D. 2012. The OAuth 2.0 Authorization Framework. IETF RFC 6749. http://tools.ietf.org/html/rfc6749.

Hildebrandt, M., & de Vries, K. (Eds.). 2013. Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology. Routledge.

Hui, Y., & Halpin, H. 2012. Collective Individuation: A New Theoretical Foundation for post-Facebook Social Networks. In AISB/IACAP World Congress 2012

Ibrahim, Yasmin. 2008. The new risk communities: Social networking sites and risk. International Journal of Media and Cultural Politics 4 (2): 245-253.

ITU. 1984. Abstract Syntax Notation One ITU-T Rec. X.680 | ISO/IEC 8824-1.

Jones, P. Salgueiro, G., Jones, M., and Smarr J. Web Finger. IETF RFC 7033. https://datatracker.ietf.org/doc/rfc7033/

Jordan K., Hauser J., and Foster, S. 2003. The augemented social network : Building Identity and Trust into the next-generation Internet. First Monday, 8(8).

Dan Kaminsky, Meredith L. Patterson, and Len Sassaman. 2010. PKI layer cake: new collision attacks against the global x.509 infrastructure. In Proceedings of the 14th international conference on Financial Cryptography and Data Security (FC'10), Radu Sion (Ed.). Springer-Verlag, Berlin, Heidelberg, 289-303.

Klyne, G. and Carroll, J. 2004. Resource Description Framework (RDF): Concepts and Abstract Syntax. Recommendation, W3C. http://www.w3.org/TR/rdf-concepts/

Kuner, C., Cate, F. H., Millard, C., & Svantesson, D. J. B. 2012. The challenge of 'big data'for data protection. International Data Privacy Law, 2(2), 47-49.

Kuss, D. J., & Griffiths, M. D. 2011. Online social networking and addiction—a review of the psychological literature. International journal of environmental research and public health, 8(9), 3528-3552.

Barbera, D. L., Paglia, F. L., & Valsavoia, R. 2009. Social network and addiction. Annual Review of Cybertherapy and Telemedicine 2009: Advanced Technologies in the Behavioral, Social and Neurosciences, 144, 33.

Langheinrich, M. 2001. Privacy by design—principles of privacy-aware ubiquitous systems. In Ubicomp 2001: Ubiquitous Computing (pp. 273-291). Springer Berlin Heidelberg

Lanier. Jaron. 2013. Who Owns the Future? Simon & Schuster, New York, NY.

Lyon, D. (Ed.). 2002. Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination. Routledge.

Lyon, D. 2007. Surveillance studies: An overview. Polity.

McDonald, A., & Cranor, L. 2010. Beliefs and behaviors: Internet users' understanding of behavioral advertising. TPRC.

Morahan-Martin, J., & Schumacher, P. 2000. Incidence and correlates of pathological Internet use among college students. Computers in Human

Nottingham, M. Web Linking. IETF RFC 5988. October 2010.http://tools.ietf.org/html/rfc5988

Oosthuyzen, M. 2012. The seductive power of Facebook, Online Us, Social Media design or Decline May 24, 2012.

Panzer, J. 2010. The Salmon Protocol. http://salmon-protocol.googlecode.com/svn/trunk/draft-panzer-salmon-00.html

Parecki, A. and Walters, B.. 2014. WebMention. https://github.com/converspace/webmention/blob/master/README.md

S. Perreault and P. Resnick. 2011. vCard Format Specification. IETF RFC 6350. http://tools.ietf.org/html/rfc6350

J. A. Pouwelse, P. Garbacki, J. Wang, A. Bakker, J. Yang, A. Iosup, D. H. J. Epema, M. Reinders, M. R. van Steen, and H. J. Sips. 2008. Tribler: a social-based peer-to-peer system: Research Articles. Concurr. Comput. : Pract. Exper. 20, 2 (February 2008), 127-138.

Prodromou, E., Vibber, B., Walker, J. and Copley, Z. 2010. OStatus. Draft Specification http://ostatus.org/specification

Reed, D. and McAlpin, D. Extensible Reousrce Identifier (XRI) Syntax V2.0. OASIS Working Draft. 2005.

Riva, G. 2005. Virtual reality in psychotherapy: review. Cyberpsychology & behavior, 8(3), 220- 230

Riva, G. 2010. I social network. Il mulino.Riva, G., Milani, L., & Gaggioli, A. (A cura di). 2010. Networked Flow: Comprendere e Sviluppare la Creatività di Rete. Milano: Edizioni LED. Online: http://www.ledonline.it/ledonline/Networked-Flow-Riva.html.

Riva, G., & Galimberti, C. (Eds.) 2001. Towards cyberpsychology: mind, cognition and society in the Internet age (Vol. 2). IOS Press.

Rodotà, S. 2012. Il diritto di avere diritti. Laterza.

Roosendaal, A. 2011. Facebook tracks and traces everyone: Like this!. Tilburg Law School Legal Studies Research Paper Series, (03).

Mayer-Schönberger, V., & Cukier, K. 2013. Big Data: A Revolution that Will Transform how

Sleevi, R. Web Cryptography API. WW3C Working Draft 2013. http://www.w3.org/TR/WebCryptoAPI/

Smarr, J. 2008. Portable Contacts 1.0, Draft specification, August 5th 2008. http://portablecontacts.net/draft-spec.html

Solove, D. J. 2006. A taxonomy of privacy. University of Pennsylvania Law Review, 477-564.

P. Saint-Andre. 2004. Extensible Messaging and Presence Protocol (XMPP): Core, IETF RFC 3920. http://www.ietf.org/rfc/rfc3920.txt

Stalder, F. 2002. Privacy is not the Antidote to Surveillance. Surveillance & Society, 1(1), 120-

Stiegler, B. 2008. Technics and Time: Disorientation. 2 (Vol. 2). Stanford University Press.

Thomke, S., & Von Hippel, E. 2002. Innovators. Harvard business review,80(4), 74-81.

Berners-Lee, T., Hendler, J., & Lassila, O. 2001. The semantic web. Scientific american, 284(5),28-37.

Berners-Lee, T., Hall, W., Hendler, J. A., O'Hara, K., Shadbolt, N., & Weitzner, D. J. 2006. A framework for web science. Foundations and trends in Web Science, 1(1), 1-130.

Van der Sloot, B., & Borgesius, F. Z. 2012. Google and Personal Data Protection. In Google and the Law (pp. 75-111). TMC Asser Press.

Virno, P. 2008. Multitude: Between innovation and negation. Semiotext.

Von Hippel, E. 1986. Lead users: a source of novel product concepts. Management science, 32(7), 791-805.

Von Hippel, E. 2005. Democratizing Innovation. MIT Press, Cambridge, MA.

Wark, M. 1999. Celebrities, Culture and Cyberspace. Pluto Press, Australia.Wagner, E. L.,

Wong, R. 2013. Proposed Data Protection Regulation 2012: Data Security Breach Notifications. In Data Security Breaches and Privacy in Europe (pp. 25-29). Springer London.

World Economic Forum. 2011. Personal Data: The Emergence of a New Asset Class

Wright, D., & De Hert, P. 2012. Introduction to privacy impact assessment(pp. 3-32). Springer Netherlands.

Wu, T. 2000. The SRP Authentication and Key Exchange System. http://tools.ietf.org/rfc/rfc2945.txt

Zimmer, M. 2008. Critical perspectives on Web 2.0. First Monday 13(3)