NEXt generation Techno-social Legal Encryption Access and Privacy nextleap.eu

Grant No. 688722. Project started 2016-01-01. Duration 36 months.

DELIVERABLE D6.6

STANDARDISATION REPORT

Harry Halpin (INRIA)

Beneficiaries:	**INRIA** (lead)
Workpackage:	**6**
Description:	The standardization-related efforts and achievements will be collected in this report, with a focus on federated identity, secure asynchronous messaging (email), synchronous messaging, and privacy- enhanced crowd-sourcing for collective intelligence via privacy-enhanced cloud analytics. The state of the standardization of each of these deliverables and their uptake by both grassroots-activists, governments, and industry will be assessed

Version:	1.0
Nature:	Report (R)
Dissemination level:	Public (PU)
Date:	2018-12-31

Project co-funded by the European Commission within the Horizon 2020 Programme.

Contents

1 Standardization Achievements of NEXTLEAP **3**

2 The Landscape of Decentralized Standards: How NEXTLEAP fits in **4**

3 Identity: The Foundation of the Social Web **5**

4 Metadata: The Semantic Web Revisited **8**

5 Transport: The Bits on the Wire **10**

6 Discovery **12**

7 Lessons Learned **14**

8 Next Steps: Blockchain Technology **15**

1 Standardisation Achievements of NEXTLEAP

Over the course of the project, NEXTLEAP has leveraged the protocols that were created within the context of the NEXTLEAP project within standards bodies where appropriate. Overall, this has met with very different kinds of success depending on the standard at hand and the kind of community – and thus governance – built around the protocol. The deliverable D3.6 features case studies of governance and standardization based on detailed ethnographic work around Autocrypt, MLS, and blockchain technologies. In contrast, this report quickly captures the highlights of the standardization of NEXTLEAP, and then situations NEXTLEAP's work within the larger legacy of failed previous attempts to decentralize the larger Social Web. Although the protocols designed by NEXTLEAP are quite low-level, we argue that the do not easily replicate the failures of the previous standardization effort as we synthesize the wide-spanning history of decentralization via standardization.

NEXTLEAP featured four main technical components: federated identity, e-mail (originally entitled "asychronous messaging" although used by Delta.chat for instant messaging), secure messaging (originally entitled "synchronous messaging" although secure messaging protocols like Signal and MLS now handle messages in the asychronous setting, just not backwards compatible with e-mail), privacy-enhancing analytics.The standardization assessments of NEXTLEAP are as follows:

- **ClaimChain (Federated Identity)**: Claimchain is currently *mature but not standardized*. The reason is because it is still a research project that exists primarily as running code (see D4.1), and currently as it is a variant of popular blockchain technology but does not belong to an existing cryptocurrency such as Bitcoin or Ethereum, there is not a clear standardization process for it. However, it has been proposed to the MLS (Message Layer Security) IETF Working Group as one method for future federation standardization, and is already as running code being integrated against Autocrypt (see D5.2). Therefore, Claimchain as such as no actual uptake from grassroots activists or governments as it has not yet been integrated into popular applications, although it does have uptake and interest from the open-source industry community around PGP and Autocrypt. Nonetheless, as there is much interest from activists and governments in MLS, the eventual standardization of ClaimChain is still possible as a core option within Autocrypt (as explored in D5.3) and eventually even MLS.
- **Autocrypt (Asynchronous Messaging)**: Autocrypt is *informally specified*. This community-led effort is still aiming primarily to fix the usability problems in PGP discovered in WP3 via the "bottom-up" community-driven strategy with key open source projects like Thunderbird and new projects like Delta.chat, and thus it was felt that the much more formal IETF standards process would exclude too many open-source community programmers. However, the effort has attracted the interest of the long-dormant OpenPGP Working Group at the IETF, and so it is expected that the cryptographic problems identified by D4.3 be adopted into the next iteration of the OpenPGP Working Group. While in 2016, there seemed almost no interest and cohesion in the OpenPGP community and thus the IETF was only willing to charter elementary changes to update cryptographic operations in OpenPGP but no substantial changes, it is to expected in that after the end of NEXTLEAP, there will be a revised OpenPGP Working Group.
- **Message Layer Security (Synchronous Messaging)**: MLS is *undergoing standardization*. This effort is being directed via a traditional "top-down" standards process at the IETF, with the Working Group and all in-progress standards being available online.[1]. After the problem was noticed by NEXTLEAP in WP3 and an initial solution (Asynchronous Ratchet Trees) proposed by University of Oxford researchers [3], the latest version from Inria (TreeKEM) is being formally verified in D4.3, and has already been made part of the standard. Although Signal has so far remained silent on whether they would join the effort, both European messaging providers such as Wire and Silicon Valley messaging providers (Google, Cisco, Facebook, and Facebook's WhatsApp messenger as well) have all endorsed the effort, as well as large open-source projects like Mozilla who plan to integrate MLS into the browser. One of the largest victories by NEXTLEAP researchers is ensuring that MLS remains decentralized by working with Wire and even gaining the support of Google for MLS being a federated protocol, as outlined in D2.3. This work will con-

[1] https://datatracker.ietf.org/wg/mls/about/

tinue to attract the interest of academics, as shown by the Workshop on Secure Messaging at Eurocrypt 2019, co-chaired by Inria.[2]

- **Privacy-Enhanced Analytics)**: The work on privacy-enhanced "wisdom of crowds" is currently *still research*. The work done by NEXTLEAP in D2.4 led to a number of innovative consensus protocols such as Blockmania and ways to detect attacks in decentralized networks such as Sybil-Quorum, as well as new results in privacy-enhanced information retrieval. This work is not immature, as it has already produced code given in D5.4 that is being used by startups and open-source projects such as ChainSpace on an industrial scale. However, the work is very new and consists of a number of separate components, and thus does not present as a single item itself as ready for standardization.

2 The Landscape of Decentralized Standards: How NEXTLEAP fits in

Far from being a mere technical concern, the promise and failures of decentralization via open standards in terms of Internet governance are a matter of pressing public concern. The fate of the Internet as a common socio-technical infrastructure for our personal data is one of the most intimate yet political questions of future generations. Due to controversies around the selling of personal data by companies like Cambridge Analytica and the passage of laws like the General Data Protection Regulation in the European Union, the general public is waking up to the world-historical danger posed by the system of control created by the rise of a few centralized platforms such as Google and Facebook.

Interestingly enough, ordinary software engineers were aware of the dangers of centralization at the very advent of social networking[3] and have long been attempting to build practical decentralized systems to counter these threats to human freedom. Given that the original World Wide Web itself was created via open standards like TCP/IP, HTTP, and URIs, it should come as no surprise that the strategy deployed by these grassroots computer programmers to counter the control of Google, Facebook, and other centralized platforms was primarily based on creating new open standards for decentralizing social data and protecting personal data.

Social data is data about the relationships between people and their environment, and so can be considered a commons in terms of ownership and possibly governance [13]. Social data would clearly include data like maps and public government or scientific data. On the other hand, personal data is considered data that reveals information about an individual, and thus the ownership and governance of this data can be considered a matter of self-determination of the individual [14]. Personal data would include personal names, addresses, identity card numbers, and so on. Of course, this division relies on inscribing a number of ontological categories that ultimately may not actually hold true. *All data is social* as it is dependent on a complex web of social relationships that characterize a process of collective individuation (as put by Stiegler), where - at any given moment - the individual is considered a result of continual co-evolution with their socio-technical infrastructure [9]. However, what is clearly self-evident is that regardless of the dialectic between social collectivity and individual autonomy, data itself is not as simple as individual or even collective property rights: One does not "own" data in the same manner one owns a coat or a house. Data around "friends" is neither clearly social nor personal: To which "friend" does the link of friendship belong? However, this does not mean that individuals have no rights over their data and so social data should automatically belong to whichever platform, such as Facebook, harvests this data. Instead, data literally co-creates the individual and society, and so their digital data is part of their very self. This viewpoint towards personal data brings it into the realm of fundamental rights, where the harvesting of personal data is more akin to a new kind of cognitive slavery, and just as everyone has the right to both self-determination in terms of their body and thought, and via mutual association vis-á-vis larger society, social data should also be controlled ultimately by the people who co-create this data. In terms of *decentralization*, no

[2] https://eurocrypt.iacr.org/2019/affiliatedevents.html

[3] Let us not forget that the first versions of Twitter actually offered the support of decentralized XMPP, and this decentralized Twitter was turned off not by the programmers, but by the management who could see no demand to support decentralization. At the time, users didnt understand decentralization, much less want it (personal communications with Blaine Cook, founding engineer at Twitter).

trusted third party should be given control of data, but instead individuals and groups should maintain control over their own data [16].

Although there have been attempts to inscribe the autonomy of data via legal means such as the General Data Protection Regulation, it is an entirely another question whether there can be decentralization of our social data via technological means like open standards. The history of this engineer-based movement for decentralization is far older than the advent of "blockchain" technology, although the advent of blockchain-based systems offer something new: An approach of guaranteeing the integrity of global common knowledge, albeit at the cost of privacy. While there is a frenzy of activity around Ethereum, Ethereum has yet to prove itself as a working technical alternative to Silicon Valley's centralization of the Web, as Ethereum's initial design ignored many of the lessons learned from computer science research into distributed systems and programming language theory. More dangerously, technically the approach to simply decentralize existing social systems may inadvertently lead honest yet naïve programmers to create a new and even more dangerous form a control society masquerading as a liberatory future: For example, a version of the Chinese 'social credit' system could easily be built in a decentralized manner using blockchain-based smart contracts.[4] While it is possible Ethereum or another as yet unnamed technology will usurp the Web, at this point the issues of governance and standards for blockchain technologies are still in their early stages. Given that the future of human autonomy itself now is intertwined with technologies, we must revisit the tangled history of failed attempts to decentralize the Web 1.0 that was recuperated into the Web 2.0, so that those that creators of the next Web do not repeat the tragedy of the Web 2.0 as a farce. It is precisely this farce that NEXTLEAP hopes to prevent by pushing for strong encryption and decentralized identity.

3 Identity: The Foundation of the Social Web

In a simplistic abstract sense, identity is the capability to distinguish one object from another.[5] This capability typically results in socially embedding discrete names on the level of an individual, such as a personal names and uniquely identifying national identity numbers. The first article that imagined the potential political promise of decentralized identity standards was *The Augmented Social Network: Building Identity and Trust into the next-generation Internet*[10], which began with a statement that seemed to be naïve, but also prefigured many of the fundamentally political questions at the heart of social networks: "Might a 'next generation Internet' help to reinvigorate democracy by providing a platform that makes it easier for citizens to inform themselves about public policy debates, self-organize, and participate in the process of governance?" The primary claim was that the main missing technical component was "a form of persistent identity that serves civil society" for the Web that would "cross traditional borders," and so create an "augmented social network," doing for the social collectivity that which Engelbarts project to augment the human intellect had done for the individual with the invention of the personal digital computer [10]. The vision of persistent identity was that "each time we go from one social network to another we do not need to restate who we are, what our interests are, or who we know" so that people would be able to re-engineer a new kind of social network, as "the design of the technical infrastructure underlying online communication is increasingly determined by for-profit entities that seek to monetize every aspect of our discourse" rather than realize the global democratic potential of the Internet [10]. Note this article on the dangers of the possible commercial centralization of identity was written before Facebook was even founded. Ironically, the authors proposing a persistent identity for decentralized social networking seemed to think the technical problems were trivial.

The Web does not include any notion of personal identity by design as the same web-page was meant to be displayed to every user in order to enable scalability via the REST architecture [6]. While web-sites were designed to have persistent identities such as *www.example.org* via the Domain Name System (DNS), users had no identifiers or personal data associated with them, although this led to the creation of invisible, ad-hoc

[4] https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion

[5] For a detailed metaphysical and cognitive treatment of identity, Brian Cantwell Smith's *On the Origin of Objects* presents a metaphysics where objects are "carved" via registration from the underlying metaphysical flux [2]

techniques for associating personal data with offline identities via third-party "tracking" cookies. Instead of a single cross-Web identity, users had to create a new identity - a set of attributes containing personal data - across every website such as a new name, a new profile photo, and a new password. In contrast, DNS was invented at nearly the beginning of the Internet itself and pre-dated the Web, and has long been a centralized registry for the identity that mapped human-readable domain names for web-sites to IP addresses.[6] Originally this first internet identity infrastructure for names created and ran by a single IETF member, Jon Postel on a voluntary basis. When Postel managed to reconfigure eight of the twelve 'root' DNS servers via a simple email (against the wishes of the US government), there was a crisis of formal decision-making inside the IETF. In response, the US government approved the creation and transfer of DNS to a California non-profit called ICANN (Internet Corporation for Assigned Names and Numbers), given a limited duration license by the US Department of Commerce, and ICANN took over the running of DNS in 1998 (shortly after Jon Postel died).[7] ICANN was given a nominally democratic structure and eventually in 2014 the United States government handed over power over DNS indefinitely to ICANN. The sheer centralization of naming authority in the hands of ICANN was viewed by many advocates of decentralization with suspicion. There were also concerns of its financial monopoly on domain names, where in a form of "digital feudalism," ICANN enables domain registrars to charge web-site owners rent for simply having a name on the Web, and it handed over the selling of top-level domains like *.com* to registrars like Verisign or to nation-states such as *.fr* for France. Nevertheless, the system remained fairly functional for decades, but ICANN never assumed power of giving digital identities to end-users.

It struck a few enterprising individuals that one possible business model would be to create a new kind of centralized DNS for the personal identity of individuals and organizations, and ordinary people would have to buy their identity from their start-up. In other words, their startup could be a for-profit self-anointed ICANN for individual identity. Although replacing ICANN seems to be progressive, these start-ups would replace a nominally democratic non-profit with a startup that would have a for-profit governance structure. The first to try this business model was Drummond Reed, whose startup Cordance created their own system of "eXtensible Resource Identifiers" (XRIs). With the new centralized XRI scheme, individuals would be given new names such as *+drummond*.[8] XRIs could be resolved by XDI (XRI Data Interchange), an equivalent to the DNS system for XRIs, which would in turn retrieve an XRDS (Extensible Resource Descriptor Sequence) that then describes the person's attributes such as age and photo. However, XDI was ran by a single organization called *XDI.org* that held a license to patents from Reed's Cordance company [11]. Perhaps XRIs were standardized at OASIS because, unlike W3C or IETF, OASIS allowed its standards to contain patents and for the license-holders to demand licensing fees. In fact, Reed's Intermind startup even brought up the patents to the W3C, forcing the W3C to create a royalty-free patent licensing scheme.[9] Although Reed claimed to give the patents to the non-profit XDI.org, others at OASIS such as Verisign, believed they had been re-licensed and the entire scheme was a get-rich-quick scheme by a patent troll. Regardless, while XRIs was under development, Reed tried to insert XRIs in as many other standards as possible. Their strategy for deployment of XRIs seemed to include getting adoption via new standards that companies would implement by default, including a newly minted OpenID standard. Eventually the W3C stepped in and OASIS closed the standardization process, effectively ending XRIs.[10] Attempting to force a get-rich-quick scheme business model into an "open standard" was rejected by the governance of the more authoritative standards bodies like the W3C, which although there is no formal governance relationship between W3C and OASIS.

The developer Brad Fitzpatrick, creator of LiveJournal, in 2005 wrote (with the help of David Recordon at SixApart) a blog post called "Thoughts on the Social Graph" where he stated "there doesn't exist a single social graph (or even multiple which interoperate) that's comprehensive and decentralized. Rather, there exists hundreds of disperse social graphs, most of dubious quality and many of them walled gardens."[11] In response, Fitzpatrick and Recordon created OpenID.[12] The vision of OpenID was originally that an OpenID would be a "Single Sign-

[6] https://datatracker.ietf.org/doc/rfc1591/
[7] https://www.wired.com/2012/10/joe-postel/
[8] https://www.oasis-open.org/committees/download.php/15376/xri-syntax-V2.0-cs.html
[9] https://www.w3.org/1999/04/P3P-PatentBackground.html
[10] https://lists.w3.org/Archives/Public/www-tag/2008May/0078.html
[11] http://bradfitz.com/social-graph-problem/
[12] https://openid.net/specs/openid-authentication-2_0.html

On" client, allowing a user to login into many different web-sites. The main issue is that they envisioned that a user would become their persistent OpenID identifier across all digital services. This OpenID identifier could be an XRI, or perhaps something more mundane like a URL. If a website supported OpenID, a user could simply sign-in *once* into their "identity provider" and then other websites could import their identity and related personal data into a website without even using their password. Sadly, OpenID made a crucial mistake: People don't confuse their own personal identity with their credit card number, and so they are equally unlikely to confuse their personal identity with a text string like a URI or OpenID. By identifying users by things they didn't understand, such as XRIs and URIs, OpenID was expecting to create an entire new mode of social interaction. Rather than try to force new modes of social identity (where users created new identifiers or cut-and-paste URIs into forms to identify themselves), decentralized social systems should build on existing social patterns such as e-mail and telephone numbers that users already understand and use on a daily basis. The fundamental mistakes made around user experience and an overly-complicated standard led to virtually no uptake, so most major sites like Facebook and Google eventually gave up on OpenID 1.0 (and 1.1) by 2015. The real value of open standards comes from patent commitments and the building of an actual community of developers around these standards, yet this community assumes users actually want these standards to begin with and can use them in their everyday lives.

OAuth is currently the most successful standard for transferring personal data between sites, but OAuth does not specify a name for an identity like XRIs or OpenID. Before OAuth, if a web application wanted to retrieve a user's personal data from a website on a social networking platform like Facebook or a large platform like Google, the password for a user's Google or Facebook account had to be transmitted to the web application, which posed a security threat as it allowed third-party applications unrestricted access to personal data, like all of a user's email. Twitter engineer Blaine Cook didn't want to have the responsibility for the passwords of his users, and realized that the OpenID architecture had the aforementioned usability concerns. Therefore, in 2006 Blaine Cook started the OAuth standard to enable the secure authorization of the transfer of information from one site to another.[13] In essence, OAuth creates a time and scope delimited token, and allows one website to request user information, and then redirects the user back to the site to authorize this access. If the user agrees, the user is redirected back to their originating website, which then gains the ability to get a scoped and permissioned access to the user's personal data via a shared secret the user has explicitly authorized. The protocol did not specify the kinds of data that could be transferred or the user experience. OAuth went through standardization at IETF (Internet Engineering Task Force), the oldest and largest multi-stakeholder standards bodies on the Internet. The IETF fixed a number of security holes, although it increased in complexity through different versions. Regardless, OAuth was by far the most successful of all the standards to decentralize social data, as Google, Facebook, and other large identity providers took up OAuth and many smaller sites enabled OAuth-backed login protocols. There are today perhaps more OAuth transactions than Visa transactions. However, OAuth identity providers became centralized due to the "NASCAR problem," namely that users could not either run their own identity provider or cognitively manage to chose from a large list of identity providers, leading personal identity to be recentralized in Google and Facebook Connect. Ironically, Google, Facebook, and Twitter all deploy a profile of OAuth called OpenID Connect.[14]

In general, Autocrypt and MLS have learned from the lessons and both support decentralized identity. Both Autocrypt and, likely MLS, support the traditional federation of domain names over the internet as *username@domain* as currently already accomplished in e-mail standards. By not trying to build an all encompassing digitial identity but already building on the affordances provided by e-mail and messaging, this identifier scheme for decentralized identity is likely to be sucessful. Of course, it does feature the issue brought about by Moxie Marlinspike that it does not natively support access to a contact list on the device, as Signal and WhatsApp access via use of the phone number. However, other applications like Wire make this option and rely on data-mining and even data-sharing on the server-side, which also features privacy risks as it requires the contact list being moved from the user's local device to the server. NEXTLEAP advocates putting the user in control by allowing the user to link contacts optionally on the mobile list on the server-side but keeping them privacy-enhanced via the use of cryptography, as can be done via the use of technology for PIR and using ClaimChains, while still main-

[13] https://tools.ietf.org/html/rfc6749

[14] https://openid.net/connect/

taining the sharing of identifiers via easy-to-remember identifiers such as *username@domain*. Furthermore, as explored in D5.3, these identifiers can be linked to Claimchains, allowing any attribute, ranging from name (or pseudonym) to favorite book, to be expressed in a privacy-enhanced and user-centric fashion. This allows both privacy-enhanced generic claims to be used by messaging and e-mail standards, and even use MLS in a purely P2P environment as would be required by blockchains where *user@domain* can be replaced by a link to a self-describing hashed value of a key.

4 Metadata: The Semantic Web Revisited

Metadata can be thought of as how to categorize everything that we may want to identify, including the attributes of the identified objects. These attributes may include data such as the favorite color and city of birth of an individual, but may also include links to other objects (such as a list of friends) or categories (such as nationality and profession). Therefore, it seems that either a tightly defined list of categories and data needs to be defined (as done by OpenID Connect or an open-ended standard way of describing all possible metadata could be used, as put forward by Tim Berners-Lee's Semantic Web standards [1], a set of standards for metadata developed by the W3C. At the time in 1999, the most widely used standard for data formats was W3C's XML (eXtensible Markup Language),[15] a generalization and simplification of SGML (Standard Generalized Markup Language).[16] This was not accidental, as SGML was a structured language for books that was the inspiration for HTML (HyperText Markup Language). However, while XML was suitable for hierarchical data, and although HTML did manage to add links (as did XML with the confusing W3C XLink standard[17]), such a language was not suitable for graph-based data. The social media revolution was conceived when the link in HTML was generalized to be more than a link that took a user between web-pages, but a link that represents friendship. In this way, the concept of a network of friends became transformed into a social graph.

As there was at the time no standard for graph-based data, the W3C decided to invent RDF (Resource Description Framework) in 1999.[18] RDF was an attempt to create a standard for decentralized information sharing, which Tim Berners-Lee assumed would more naturally fit on top of the link-based Web than the tree-based XML standard. Just as Bitcoin came into prominence as a technique to get around the financial blockade of Wikileaks, the Semantic Web's utility for describing social relationships had a real use-case due to repression and censorship on the Internet. The earliest known decentralized social network, the Friend-of-a-Friend (FOAF) network, came into being as an attempt to build a social network that could not be censored by the Iranian government, who at the time in 2005 had cut Iranian users off from the Internet.[19] FOAF was not only the first vocabulary for a decentralized social network, but came to be in 2000 before centralized social networks such as Facebook (2004) and Myspace (2003) and even Friendster (2002) were even founded.[20]

The W3C Semantic Web arrived stillborn as RDF made a number of design errors. First, although a more user-friendly syntax called N3[21] that appeared similar to JSON was proposed by Tim Berners-Lee, instead the W3C standardized a very difficult syntax, RDF/XML, that attempted to squeeze the RDF graph model into an XML tree-based serialization (as explained in Section 5).[22] The W3C justified the impossible-to-read syntax of RDF/XML by asserting that only machines would read RDF via parsers and programmers would use some more advanced visualization tools, but twenty years later these tools have not appeared. The second mistake made by RDF was its believe that simplistic logic-based inference was necessary to build into the syntax so that RDF could detect, for example, that a "friend" was a type of "social connection" and the inverse of "enemy." The

[15] https://www.w3.org/TR/xml/

[16] https://www.iso.org/committee/45374.html

[17] https://www.w3.org/TR/xlink/

[18] https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/

[19] http://xmlns.com/foaf/spec/. Also see the article *Open Social Networks: Bring Back Iran* by Dan Brickley, inventor of FOAF: http://danbri.org/words/2008/01/07/249

[20] It should be noted that the first social networking sites can be considered AOL Messenger and SixDegrees, which were founded in 1996, before FOAF but also before well-known social networking sites like Myspace and Facebook.

[21] https://www.w3.org/DesignIssues/Notation3.html

[22] https://www.w3.org/TR/rdf-syntax-grammar/

lead semanticist of knowledge representation-based artificial intelligence, Patrick Hayes, did manage to make a fairly straightforward formal semantics to enable these kinds of inferences.[23] Still, the very term "semantics" caused much confusion but very little in the way of working applications. Indeed, academics took over the W3Cs standardization of the Semantic Web. With a plethora of Semantic Web standards ranging from the RDF query language SPARQL, to more than half-a-dozen mostly incompatible versions of the Web Ontology Language OWL, and even attempts to interoperate with XML via GRDDL, chaos reigned. The entire Semantic Web stack of technologies to this day are difficult to use and inefficient for real-world deployment compared to traditional databases, and attempts to add "Web Services" to the Semantic Web failed.[24] Attempting to force adoption of a technology via premature standardization is a recipe for failure.

Another attempt to maintain the spirit of the decentralized Web was microformats, an initiative founded by a group of developers around Tantek Çelik.[25] Microformats was created purposefully outside of the W3C as it was felt that the W3C would both bureaucratically slow down development and also attempt to force the use of RDF. The idea was quite simple, that interoperable "microformats" would be embedded in HTML that would allow the uniform sharing of data across web-sites. Technically, if the right semantic tags (i.e. tags with a meaning, not for presentation) were embedded into HTML markup using the `span` and `div` tags, then automated web-scrapers could extract metadata from the website's HTML itself. Microformats was a design argument against RDF and RSS, which wanted websites to host their metadata in separate files from the HTML itself. The argument used by microformat supporters was that websites were incentivized to keep their HTML up-to-date, but not serve separate files for metadata. As it was easy to add microformats to existing HTML, microformats did indeed spread like wildfire across the Web in 2007, with adoption by Web 2.0 sites such as Flickr. However, ironically as search engines such as Yahoo! SearchMonkey and Googles RichSnippets started consuming large amount of microformats, it became increasingly obvious that web-developers were very error-prone when adding microformats to their websites [12].

Despite the large amount of developer interest in microformats and academic interest in the Semantic Web, ultimately both of these initiatives were recuperated by the Facebook and Google platforms. While the W3C tried to catch up to make microformats compatible with the Semantic Web via the awkwardly-named GRDDL standard to convert microformats to RDF[26] and the RDFa standard[27] to allow RDF to be embedded directly into web-pages, Facebook had its own plan to embed metadata into the Web for its own purposes. David Recordon, known for being the co-editor of "Thoughts on the Social Graph" and designer of OpenID Connect, joined Facebook. At Facebook, Recordon made a new specification, the *Open Graph Protocol*, to embed a limited number of vocabulary items in web-pages.[28] This Open Graph Protocol could describe strictly delimited types of books, movies, food, and other items of interest, so the Open Graph Protocol was neither "open" nor a RDF "graph" nor even a "protocol." In a self-serving clever twist by Facebook, it also came with cut-and-paste Javascript that would allow any web-page to embed a "Like" Button that would harvest this metadata and send it back to Facebook. This effort took off, and as of 2017 over 6% of the top 10,000 sites on the Web features a "Like" Button.[29] Ironically, the Semantic Web's largest deployment to date is Facebook's "Like" button [12], allowing Facebook to collect metadata on user likes across the entire Web in a privacy-invasive manner.

Google was not to be left out of the proprietary metadata game. While the W3C believed that RDF would enable the creation of hand-crafted decentralized ontologies describing domain-specific metadata formats, in reality Google ended up centralizing the creation of metadata.[30] An evolution from the parsing of microformats by Google Search and Google's own competitor to RDF, microdata, that they put into HTML,[31] *schema.org* systematized the kinds of domains that Google users were trying to discover, ranging from shopping to recipes.

[23] https://www.w3.org/TR/rdf-mt/

[24] While there were entire books published and billions of euros spent in European Commission project grants, there is to date no working Semantic Web Services. For the details of perhaps the largest failed research attempt of the Web, see Dieter Fensel et al.[4].

[25] http://microformats.org/

[26] https://www.w3.org/TR/2007/REC-grddl-20070911/

[27] https://www.w3.org/TR/2015/REC-rdfa-core-20150317/

[28] http://opengraphprotocol.org/

[29] https://trends.builtwith.com/cdn/Facebook-CDN

[30] https://schema.org

[31] https://www.w3.org/TR/microdata/

Google refused to work with the W3C, instead opting to standardize schemas with other browser vendors such as Microsoft and Yandex, and eventually letting the community provide input via a non-binding W3C Schema.org Community Group. In the end, Google incorporated metadata from Wikipedia's Wikidata effort [17], creating the Google Knowledge Graph. The Google Knowledge Graph is a closed and proprietary version of the Semantic Web that serves as the backbone of Google's massive data collection efforts and powers their machine-learning algorithms. Rather than open standards for the decentralized social data, each Silicon Valley company has their own knowledge graph, a closed proprietary metadata collection. RDF currently is used for open public data and some other fields like library science, today we still do not have the ability to specify in a standardized way social metadata. The problem may ultimately itself may not be technical: The ability to describe and represent the objects in our world in a digital space without falling into the trap of pre-inscribing all possible categories remains one of the greatest unsolved problems in the terms of metaphysics itself.

Metadata allows programmers to attempt to be amateur sociologists and philosophers; metadata is an excuse for an infinite "bikeshedding," that breeds endless discussions and little in the success of way of standardizing. Even the attempt to formalize only systems for characterizing metadata, such as the Semantic Web, tend to lead to overly restrictive schemes that do more to imprison the world than to lead to better communication. Therefore, both Autocrypt and MLS have avoided connecting any standardized metadata scheme to identity, as well as enforce any attempt to embed particular metadata schemes. Only the minimal metadata needed to discover, as explored in Sectin 6, the versions of the protocol and establish communication are embedded in the specification, as is done via SMTP headers in Autocrypt and via suggested pre-set messages in MLS.

5 Transport: The Bits on the Wire

Transport is the application-level format used to actually transfer data (and metadata) from one identity to another over a network protocol (such as TCP/IP). Traditionally on the Web the transport mechanism was given by HTTP. In terms of decentralization, RSS was one of the most successful open standards that nearly - but ultimately failed - to decentralize the Web. RSS originally was an abbreviation for RDF Site Summary in 1999, created by R.V. Guha, co-designer of the RDF standard and AI expert. However, the original version of RSS was difficult for developers to understand and even parse, due to the use of the syntax of RDF in XML, although the RSS-dev Working Group continued to evolve the design into RSS 1.0, but RSS 1.0 remained virtually unusable (which was the version of RSS championed by Aaron Swartz).[32] Therefore, a new version of RSS was created by David Winer, which succeeded insofar as he removed any traces of RDF, and stuck to a simple XML-based syntax. This version of RSS was rebranded "Really Simple Syndication" in RSS .91 and .92 to prevent confusion with the RDF version of RSS.[33] The RDF version of RSS was a failure despite the informal Working Group continuing to work till 2008. The simplified XML version of RSS allowed blog rolls, audio files, and other data to be syndicated across web-sites, exploding in usage from 2002 onwards. Winer's RSS played a crucial, if mostly hidden, role in the infrastructure of what was called "the Web 2.0." The spread of RSS was eventually stopped by the splintering of the standard itself, as developers and users were stymied by incompatible versions. The W3C kept attempting to press adoption an RDF-based version of RSS 1.0 due to Tim Berners-Lees desire to keep pushing the adoption of the Semantic Web via open standards. Likewise, David Winer placed the stewardship of his competing XML version of RSS, now re-branded RSS 2.0 (in order to leap-frog across RSS 1.0) in the Berkman Center at Harvard rather than W3C/MIT, as Harvard did not push for the support of RDF like the W3C.[34] Therefore, developers could not sensibly determine how to support the RSS-based decentralized web with multiple versions of distinct incompatible standards with competing version numbering schemes. Due to this ego-driven standardization failure, the decentralized RSS-based Web 2.0 was cripplied at birth. The IETF finally managed to fix the wreckage of the three different incompatible versions of RSS by creating the XML-based Atom, but by then it was too little, too late.[35] In the wild, RSS usage was split between the different incompatible

[32] RSS-dev Working Group RDF Site Summary (RSS) 1.0 2000. `http://web.resource.org/rss/1.0/spec`
[33] RSS 0.2 (2002) `http://backend.userland.com/rss092`
[34] RSS 2.0 (2003) `https://cyber.harvard.edu/rss/rss.html`
[35] `https://tools.ietf.org/html/rfc5023`

formats.[36] Facebook and Twitter dropped RSS support, and eventually in 2013 Google canceled support of their popular RSS reader. The hope of decentralizing the Web 2.0 via decentralized status feeds was dead.

The Extensible Messaging and Presence Protocol (XMPP) was a standard meant to enable real-time XML-based messaging.[37] Unlike the HTTP-based Web that required users to "pull" web-pages to their browsers, XMPP was built on a "push" model that let new content be sent to users. In addition to its core architecture, XMPP also had its own persistent and federated identity system for users, and so was a complete system for instant messaging. Therefore, XMPP had a moment of surprising success, with many chat clients adopting XMPP as a foundation of decentralized interoperability, including Google Talk. However, XMPP failed to evolve into a decentralized real-time alternative to the Web. At first this may be surprising, as new functionality, such as that needed to replicate the features of Facebook, could be added to XMPP as it was an extensible standard. Extensibility was both a blessing and a curse, as it also led to overwhelming complexity: XMPP spawned its own mini-standards body, the XMPP Foundation, wherein hundreds of extra features were added. The XMPP core became more and more unwieldy itself, eventually reaching over 200 pages; the ability of developers to implement these standards, much less make them interoperate, became non-existent. As it became more and more difficult to gain interoperability, the XMPP standards became more of an hindrance than a boon for the creation of a decentralized social web, with the lack of interoperability holding back development. One by one, client support for XMPP dwindled. In 2015, Google Chat finally completely dropped XMPP, and chat clients such as Signal and WhatsApp came about that didn't support XMPP. Indeed, attempting to replicate the organic functionality of centralized silos using a single standards-based framework led to complexity, which created a lack of interoperable implementations, causing the decline of the XMPP eco-system. What appeared to be one foundation of a decentralized Webs transport layer ended up being abandoned in 2018,[38] although it maintains some usage amongst developers and activists due to support of the OTR[39] and OMEMO encrypted chat applications.[40]

Pubsubhubbub was an invention of decentralized web pioneer Brad Fitzpatrick, who left Livejournal to work at Google and who had authored earlier the "Thoughts on the Social Graph." Pubsubhubbub standardized the "publish-subscribe" push model over HTTP, so that HTTP replace XMPP.[41] Pubsubhubbub allowed RSS-based sites to be pushed dynamic real-time content, and so could have - at least in theory - enabled the real-time updates needed for a decentralized Facebook and Twitter without relying on a clunky XMPP-based architecture. The various parts were knitted together into OStatus,[42] including a new Atom-based ActivityStreams format[43] that was meant to provide social updates such as "friend requests" adds in a decentralized manner, using Pubsubhubbub to communicate. OStatus was incarnated as a federated free software alternative to Twitter, *status.net*, eventually being given over to the Free Software Foundation as GNU Social. Yet by the time the federated social web was ready to be used, it was too late: Facebook and Twitter were entrenched, and Google had given up on attempting to produce an open social web, instead centralizing in Google Plus and shutting down their RSS reader. The W3C convened a W3C Social Web Working Group in 2015[44] to attempt to standardize a decentralized social web led by IBM, but arguments over the transport layer ended up causing the W3C to embarrassingly create *three* incompatible formats, primarily a JSON-based ActivityStreams 2.0 [15] and an competing RDF version of called Linked Data Notifications,[45] (as well as a new version of Pubsubhubbub called *WebSub*.[46] Embarrassingly, internecine wars over formats in standards bodies that have haunted the decentralized web since its inception ended up making the results of the W3C Social Web Working Group unusable two decades later in 2018.

[36] The vast majority using RSS 2.0, followed by Atom, and then previous RSS versions in 2018.

[37] https://xmpp.org/rfcs/rfc3920.html

[38] XMPP also was the backbone for the ill-fated and confusing Google Wave, which was dropped by Google in 2010 although idealistic software developers such as Kune and SwellRT are working on trying to build a decentralized social web on top of XMPP.

[39] https://xmpp.org/extensions/xep-0364.html

[40] https://xmpp.org/extensions/xep-0384.html

[41] http://pubsubhubbub.github.io/PubSubHubbub/pubsubhubbub-core-0.3.html

[42] https://www.w3.org/community/ostatus/wiki/images/9/93/OStatus_1.0_Draft_2.pdf

[43] http://activitystrea.ms/specs/atom/1.0/

[44] https://www.w3.org/Social/

[45] https://www.w3.org/TR/ldn/

[46] https://www.w3.org/TR/websub/

The key error of all these attempts to federate was to leave out encryption from the standard. It is precisely this lack of security that led original standards such as SMTP and even XMPP to be trivial targets for mass surveillance, and attempts to "bolt-on" encryption afterwards like OpenPGP for SMTP and OTR for XMPP often lead to errors in encryption, as shown by D4.3. These errors are almost to be expected, as it is difficult to make an insecure protocol secure. Therefore, both Autocrypt and MLS attempt to make end-to-end user-centric encryption the default. As a new protocol, in MLS this is simple as MLS specifies in detail the encyption and leaves the other transport layer mechanisms under-specified. This is viewed as an advantage as it allows both rich JSON-based transport encoding, but could also support compressed byte serializatons for Internet of Things and mobile devices. On the other hand, Autocrypt is left with the default insecure but widely adopted SMTP e-mail standard for transport, but uses the headers in an attempt to enable end-to-end encryption as soon as possible.

6 Discovery

Discovery involves "discovering" what capabilities that allow interaction with an identity. For telephone numbers and government identities, this is solved by centralized databases, and for URLs by a mixture of domain name servers methods, including negotiating what kinds of applications can work with a given identity. One example is the difference between a domain name as identity in DNS and the protocols that an identity supports. The same name in DNS may support two different protocols, such as *ftp://example.org* shows a name supports File Transport Protocol while *http://example.org* indicates the identity supports HTTP (HyperText Transfer Protocol). HTTP was the killer application for domain names, a protocol initiated by Roy Fielding and Tim Berners-Lee as a stateless protocol for transferring hypertext[5] A discovery protocol may be more complex, and could allow multiple types or even an extensible number of capabilities. For example, a single identity may be able to communicate via multiple protocols, such as a domain name can map to IPv4 IP addresses using the `A` record, IPv6 addresses using the `AAAA` record, and even for e-mail via the `MX` record. This is similar to a single Bitcoin address supports both Bitcoin and Bitcoin Cash. One of the goals of many identity systems that have come along after DNS is to allow the kinds of records to be more easily extensible in order to support an open-ended number of protocols and capabilities.

It was not only the overly complex RDF as technical infrastructure for social data that doomed the quest for interoperable social networks and paved the way for the centralization of our social data in the hands of a few major corporations like Facebook. The primary failure was one of human interaction. The social network designed by FOAF involved humans making a list of your friends in a plaintext document, perhaps including a rating for how much the author like them, and uploading this data to the Web for public consumption.[47]. Thus, Facebook was victorious over FOAF not because Facebook was centralized, but because Facebook created human-usable interaction, such as browsing photos and chatting, and Facebook offered a degree of privacy via access control only your friends could see your other friends! that was missing from portable and decentralized social networking. RDF was designed for open data, and never developed a standardized model of privacy or access control suitable for real-world use or enforcement of data protection laws. In order to develop these rich kinds of interaction in a decentralized manner, it was necessary to build APIs that could power decentralize social applications. Yet in order for the APIs to work, they needed to understand what capabilities were afforded by each of the nodes in a decentralized social network. Attempts such as XRDS (eXtensible Resource Descriptor Sequence) fulfilled this role in the XRI and OpenID eco-systems, and failed as those efforts collapsed.

Discovering social capabilities of identities ended up being more difficult than the capabilities of a web-page. While HTTP has a very limited set of operations, such as GET and POST, and that these operations described all possible web applications anyone could ever build, there are nearly infinite social capabilities. As the blogosphere failed to standardize around RSS due to over-sized egos and XMPP continued to decline in usage, more

[47] Obviously, for one of its original use-cases as a portable social network for activists being repressed, the storage of friendship networks in FOAF would be a potential honey-pot for the repressive government: The censor could just download everyones FOAF files to compile a list of who to imprison!

and more ordinary people went to social networking sites like Myspace, Friendster, and eventually Facebook. Google was caught by surprise by Facebook, and started to adopt a strategy of supporting open standards, such as Google Reader being based on RSS and Google Chat adopting XMPP. However, these data-based standards did not support the rich functionality of Facebook as there was no way to discover what clients supported what capabilities. As put by Moxie Marlinspike, "If XMPP is so extensible, why havent those extensions quickly brought it up to speed with the modern world? Like any federated protocol, extensions dont mean much unless everyone applies them, and thats an almost impossible task in a truly federated landscape...the implications of that are severe, because someones choice to use an XMPP client or server that doesnt support video or some other arbitrary feature doesnt only effect them, it effects everyone who tries to communicate with them."[48] Attempts to create interoperable decentralized social networks based on XMPP such as BuddyCloud uniformly failed.

In desperation, Google created OpenSocial, an API to embed social functionality across its varied early social networking web-sites like Orkut and Friendster via the OpenSocial Foundation, attempted to make OpenSocial an "simple, standards-based technologies" that other websites could use, despite not having a standards body like the IETF or W3C behind it.[49] At first, it seemed that Googles approach of creating its own standard body could compete the closed silos of Facebook. Yet ultimately OpenSocial failed to be adopted. Facebook, Myspace, and the others provided social functionality that was carefully crafted around their users. In contrast, OpenSocial allowed social functionality to be enabled on any website, but OpenSocial provided a decontextualized sociality while also remaining controlled by Google. Without the correct social context, OpenSocial simply didnt make sense to many Web developers, as it was on a confusing Java-based model. Without real open governance of the standard, it developed a poor reputation amongst advocates for decentralization for being dominated by Google. The only adopters were the world of enterprise software where unnecessary complexity is considered a virtue rather than a vice. Eventually, IBM took over the remaining shell of a standard from Google and donated it to the W3C to help form the W3C Social Web Working Group, which proceeded to fail to update the standard, instead producing incompatible RDF and JSON-based formats for federation.

Luckily, the IETF moved into the space of capability discovery with a very simple addition: *well-known* URIs. This URIs signaled for a common URI suffix for "the discovery of policy or other information about a host" such as `https://example.org/.well-known/`.[50] The exact kinds and encoding of these capabilities to be discovered at *well-known* were kept intentionally open-ended, just as OAuth did not define the kinds of parameters that could be transferred in security tokens. This technique for discovery became widespread on a per-standard basis, taken up by OpenID Connect, even simpler attempts to discover social profile data like *WebFinger*,[51] and even the failed "Do Not Track" standard by W3C.[52] Ironically, s each usage of well-known URIs was application-specific, it never could form the basis for an interoperable decentralized Social Web, and no API for rich social interactions ever was built on top of this particular discovery standard. Therefore, the social web failed to provide a coherent development environment, and so developers moved to closed and proprietary platforms such as Facebook, or built on the contact information provided by the address-book of the mobile phone in the Android or iOS eco-systems.

Discovery is still a difficult problem facing Autocrypt and MLS, and not yet explicitly addressed in MLS (although there is an initial draft put forward based on work done in D2.4) and is done via headers in Autocrypt. As explained in Section 4, the only use of metadata used by Autocrypt and MLS is to discover what version is discovered, and in the case of Autocrypt if Autocrypt is even supported at all. This minimalism also allows flexibility: In combination with Claimchain, which as explained in Section 3 allows any claim to be expressed, also allows any kind of discovery information to be expressed, such as other applications that a user supports, location of a social network profile, and so on. The future of discovery should be both private and user-centric, not in the hands of URLs controlled by centralized service providers.

[48] `https://signal.org/blog/the-ecosystem-is-moving/`
[49] `https://googlepress.blogspot.com/2007/11/google-launches-opensocial-to-spread_01.html`
[50] `https://tools.ietf.org/html/rfc5785`
[51] `https://tools.ietf.org/html/rfc7033`
[52] `https://www.w3.org/TR/tracking-dnt/`

7 Lessons Learned

Far from naïvely programming a dystopia of centralized personal collection, ordinary programmers were the first to take seriously the threats posed by Facebook, Google, and the like to our digital social lives. Before even the advent of Facebook and competing platforms like Google Plus and Twitter, programmers started building protocols to help citizens re-seize control over their personal data in order to build a decentralized and democratic social web. Yet this task ended in failure, and as of 2018 the consolidation of power and control over the social web by a few large corporations seems unparalleled. The difference is today that ordinary people are now aware of the dangers of a centralized social web due to increasing hacking attacks on these irresistible honey-pots and the abuse of personal data for political purposes. Yet given programmers had over ten years to address the centralization of social data, why did these efforts fail? To a large extent, the key failing is that the programmers tried to solve the problem of centralization via the purely technical means of standards rather than taking into account the larger social, economic, and political world into which their code was embedded.

Given a technical standards-based approach, a mundane reason for the failure of the decentralization of the Web was a failure of a unified strategy pushed by responsible standards body, due to a lack of intellectual clarity over the necessary minimal components to standardize and a simple way for programmers to implement them. This would normally be the job of a standards body such as W3C or IETF to plan, but it seemed that these standards bodies developed a strategy far too late. Up until W3C formulated a (failed) plan in 2013,[53] decentralized social standards were often ran by a small group or even a lone coder, such as David Winer's version of RSS hosted at Harvard Berkman (apparently put there to spite W3C at MIT, who would have attempted to force the usage of RDF in RSS). There were even worse iterations of this, such as the attempt to create "standards" such as XRIs by rather questionable entrepreneurs or via "one shot" standards bodies such as the OpenID Foundation. Early on, large projects such as Project Higgins of the Berkman Harvard Center and XRI-based startups like Cordance also all ended in failure, defeated by their own needlessly complex architectures. Of the plethora of standards for a decentralized social web, in terms of real-world deployment what happened (again and again) is that a few of the larger companies such as Google or IBM would adopt components as suited their business strategy (and killed, like RSS or XMPP by Google, as soon as convenient), while the rest of the components were relegated to small open source projects.

Rather than blame Silicon Valley entirely for the failure of a decentralized web, history shows that hackers are their own worst enemy. Rather than traditional multi-stakeholder standards bodies taking responsibility for developing a single suite of decentralized standards that would serve the needs of stakeholders like the general public and entrepreneurs, standards bodies transformed into strange religions around data formats. The example par excellance is the Semantic Web effort of Tim Berners-Lee and the W3C. For example, the use of RDF needlessly fractured the RSS standard and for years forced development of social standards outside the W3C. When the Social Web community decided to develop a W3C Social Web Working Group in 2014, the group failed to produce a unified standard.[54] Under the incompetent leadership of chair Arnaud LeHors of IBM,[55] the W3C Social Web Working Group produced *three* incompatible versions of the same standards for transport and metadata: The Semantic Web-centric Linked Data Notifications,[56] the microformat-enabled (rebranded "IndieWeb") WebMention,[57] and a JSON-format for ActivityStreams 2.0.[58] The reason for this train wreck of a standards Working Group was because a small fanatical cult around Berners-Lee and his Social Linked Data[59] project pushed the use of RDF, although Berners-Lee himself did not directly interact with the standards process. Afraid of offending RDF developers, the W3C pushed through RDF and IBM sent ActivityStreams 2.0 off the rails via

[53] The W3C Social Activity's scope is `https://www.w3.org/Social/`. Note that I organized the strategy and wrote the W3C Social Web Working Group charter.

[54] As founder of the W3C Working Group, I stepped down when it became clear the W3C started to force RDF on the Working Group against the will of developers.

[55] IBM seemed interested primarily in placing any patents related to the OpenSocial into W3C's Royalty-Free patent policy.

[56] `https://www.w3.org/TR/ldn/`

[57] `https://www.w3.org/TR/webmention/`

[58] `https://www.w3.org/TR/activitystreams-core/`

[59] Called "Solid," see `https://solid.mit.edu/`

starting the absurd task of creating a meta-model for all possible social actions,[60] Çelik's group of microformat developers created Micropub,[61] and the entire situation became so confused that the W3C had to publish a guide to their non-interoperable protocols.[62] Although a next-generation Pubsubhubbub simplified as WebSub[63] and ActivityPub[64] show promise (being used on decentralized Twitter clone Mastodon), in retrospect rather than co-operate in order to bring decentralization forward, engineers preferred to engage in ideological debates over data formats whose only real-world impact was preventing a decentralized Social Web from ever being launched.

NEXTLEAP has learned that premature optimization is the root of all standardization failure, and so has taken two dramatically different approaches. As the W3C and Tim Berners-Lee damaged their reputation by backing Silicon Valley in their moves to standardize DRM at the W3C, the best standards body left for decentralization seemed to be the IETF as pioneered in the IETF TLS 1.3 standard [7]. This is because the IETF has resolutely stated that "Pervasive Monitoring is an Attack" and, unlike other standards bodies, is now committed to working through security with academics, including using formal verification.[65] In the case of MLS, the IETF was seen as a natural successor to the work fixing TLS given in TLS 1.3, and the same body is also clearly the home for any e-mail related standardization resulting out of Autocrypt. Therefore, the IETF gives a clear governance and process to standardization, one that is open to anyone who can communicate.

8 Next Steps: Blockchain Technology

Could blockchain technologies create a decentralized web? When Bitcoin emerged in 2008, it seemed only useful in transferring money. Yet the underlying technology, a blockchain, seemed to offer a new technology for decentralization. While previous attempts to decentralize social data focused on incremental fixes to the Web, blockchain technologies presented a radical alternative that hearkened back to the golden days of peer-to-peer systems in 2001, but combined peer-to-peer technology with strong cryptography: Blockchain technology provides a publicly transparent and decentralized ledger that is configured to track and store digital transactions in a publicly verifiable, secure, and hardened manner to prevent tampering or revision. In essence, transactions or completed blocks in the blockchain are recorded and added to a chain in chronological order to allow market participants to keep track of transactions without central record keeping. Each node in the system gets a copy of the blockchain to maintain an immutable decentralized ledger. It is both pedagogical and necessary to take into context the failures of the original attempt to decentralize the original Web in the context of the rise of blockchain technologies. Protocols based on distributed ledgers are branded "Web 3.0" to contrast this wave with the centralized social media platforms of "Web 2.0."

As Bitcoin appeared *deux ex machina*, one crucial advantage Bitcoin had over other technical efforts is their lack of a cult-like leader, as the identity of Nakamoto has never been revealed. This is not the case with Vitalik Buterin and Ethereum, where Buterin's influence has been decisively important in issues like the crisis caused by the DAO hack.[66] Yet his influence and the influence of the Ethereum Foundation in general seems to be tempered over time by other developers. An informal governance body, often consisting of coders that have commit access to the production version of the code, has informally developed in both Bitcoin and Ethereum; a meritocracy that mirrors the early stages of the development of the IETF. Attempts to create more structured organizations such as the Bitcoin Foundation have failed, and instead a process more akin to the IETF RFC process was started by the Bitcoin Improvement Proposal (BIP) process[67] started by Amir Taaki, and then in the Ethereum Improvement Proposals (EIP) process.[68] The phenomena of "forking" is another genuine innovation that can measure popular

[60] https://www.w3.org/TR/activitystreams-core/
[61] https://www.w3.org/TR/micropub
[62] https://www.w3.org/TR/social-web-protocols/
[63] https://www.w3.org/TR/websub/
[64] https://www.w3.org/TR/activitypub/
[65] https://datatracker.ietf.org/doc/rfc7258/
[66] https://www.wired.com/2016/06/50-million-hack-just-showed-dao-human/
[67] https://github.com/bitcoin/bips
[68] https://eips.ethereum.org/

support of a standard. Popular support for these proposals can be judged simply by the amount of miners that take up the new proposal, and the more diverse eco-system of exchanges and third-party applications that follow a particular "hard fork" can empirically measure its success.

The key contribution of blockchain technologies is their trail-blazing built-in economic model based on decentralized payments. It is precisely the lack of a funding model that led many open-source and free software projects to either fail or be controlled by large platforms that are funded via surveillance-based advertising; a prime example is the sponsorship of Google of various open-source projects that are then placed behind the cloud. Yet with a new funding model and venture-capital putting billions of dollars into startups working on decentralized protocols, it appears the initial funding boom that built the "Web 1.0" in the 1990s has returned. Although this amount of funding has attracted a large amount of disreputable "get rich quick" start-ups, nonetheless a large amount of talented engineers would rather work on blockchain technologies than for Silicon Valley.

In terms of open standards, all the traditional multi-stakeholder open standards bodies were also dependent on the support and funding of a few Silicon Valley companies, as shown by the standardization of DRM led by Google, Microsoft, and Netflix at the W3C [7]. In effect, the decentralized standards of Web 2.0 failed in part due to being reliant on the noblesse oblige of a few large companies. As attempts to standardize blockchain technologies at the W3C and IETF have either so far failed or produced standards of dubious technical value, there is now the chance to re-start standardization with a fresh slate. It may be too late to save the W3C, as the Verifiable Claims work at W3C[69] is a confused mixture of Semantic Web technology and blockchain technology that ignores the fundamental fact that the Semantic Web is dependent on a centralized domain name system [8]. Blockchain advocates should beware working with the dubious characters of failed decentralization efforts in the past, as the W3C Verifiable Claims work is now backed by Drummond Reed, whose new Evernym and Sovrin Foundation seem to simply be repeating XRIs with blockchain technology. This new decentralized "identity" blockchain solution is aiming to work with IBM's centralized "blockchain solution" Hyperledger, which is also led by the same person, Arnaud LeHors, that attempted to force the use of RDF and so failed as chair of the W3C Social Web effort.

One important missing component of the decentralized Social Web was popular support. Popular support in general tends to be measured by users, but given the network effects of Facebook and Twitter, decentralized competitors such as Diaspora and Mastodon have orders of magnitudes less users. Although early pre-Twitter and pre-Facebook radical social media site Indymedia was decentralized via its use of RSS and had large usage by social movements across the world, currently social movements rely on Twitter and Facebook to reach the general public. Attempts by the European Commission to promote a decentralized social web for commons-based data using OAuth and other standards by the D-CENT project[70] also failed to gain any traction with local populations. However, one possible proxy of public support is government support. Although for the first decade of the social web, rights over personal data and privacy were signed around by terms of service, with the European General Data Protection Regulation, governments are starting to enforce strict rules on large centralized platforms. While terms of service would "promise" to respect privacy while giving the user no guarantees or procedural rights over the data, as exemplified by the mockery given to the start-ups such as Reed's failed Connect.me that advertised "We believe privacy, control, and portability are requirements, not features" while not offering a privacy policy for their collection of user data,[71] the General Data Protection Regulation will fine services for not providing basic rights and meaningful consent to users. However, rather than create decentralized alternatives to centralized platforms that support human autonomy, the General Data Protection Regulation merely attempts to fine centralized platforms for misbehavior, the regulation itself adopts a centralized paradigm of data controller and processor which does not clearly apply to decentralized systems and may even, by accident, make them illegal. So just as technical solutions without political context will not succeed in preventing centralization, political solutions that lack an understanding of the space of technical designs are equally wrong-headed.

Although political and social forces have been arrayed against the hackers working on creating a decentralized

[69] https://www.w3.org/TR/verifiable-claims-data-model/

[70] https://dcentproject.eu

[71] https://nakedsecurity.sophos.com/2011/03/10/connect-me-controversy-continues-have-your-say/

social web, it is the hackers that have defeated themselves so far. Therefore, ironically the attempts to create a decentralized Social Web have almost entirely been recuperated. The Semantic Web fueled proprietary knowledge graphs, and Berners-Lee's vision of the web as a database of open knowledge based on RDF failed to materialize. By the time OpenID had matured into OAuth 2.0, it was extensively deployed by both Google and Facebook as identity providers - as well as most web-sites and mobile apps as relying parties - to centralize control of the authentication and authorization process in the hands of a few Silicon Valley companies. The largest user of RDF and metadata ended up being the Facebook "Like" Button. NEXTLEAP has advocated the adoption of standardization based on the IETF model and formal verification in the blockchain community, as well as increasing their commitment to democracy and being wary of charismatic leaders.[72] Will the blockchain revolution bring a new decentralized web into existence, or simply become the technical infrastructure of further control and centralization? Only time will tell, but the future of decentralization is at stake, and so human freedom in an increasingly digital world depends on the new ways of governance - or perhaps better phrased, *ungovernance* - that are being developed by the next generation of "digital native" blockchain hackers.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific american*, 284(5):34–43, 2001.

[2] B. Cantwell Smith. On the origin of objects, 1996.

[3] K. Cohn-Gordon, C. Cremers, L. Garratt, J. Millican, and K. Milner. On ends-to-ends encryption: Asynchronous group messaging with strong security guarantees. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1802–1819. ACM, 2018.

[4] D. Fensel, F. M. Facca, E. Simperl, and I. Toma. *Semantic web services*. Springer Science & Business Media, 2011.

[5] R. Fielding. Architectural styles and the design of network-based software architectures, 2000.

[6] R. T. Fielding and R. N. Taylor. Principled design of the modern web architecture. *ACM Transactions on Internet Technology (TOIT)*, 2(2):115–150, 2002.

[7] H. Halpin. The crisis of standardizing DRM: The case of W3C Encrypted Media Extensions. In *International Conference on Security, Privacy, and Applied Cryptography Engineering*, pages 10–29. Springer, 2017.

[8] H. Halpin. Semantic Insecurity: Security and the Semantic Web. In *Society, Privacy and the Semantic Web-Policy and Technology (PrivOn 2017)*, 2017.

[9] Y. Hui and H. Halpin. Collective individuation: The future of the social web. *The Unlike Us Reader*, pages 103–116, 2013.

[10] K. Jordan, J. Hauser, and S. Foster. The Augmented Social Network: Building identity and trust into the next-generation Internet. *First Monday*, 8(8), 2003.

[11] D. Lim. *Patent misuse and antitrust law: Empirical, doctrinal and policy perspectives*. Edward Elgar Publishing, 2013.

[12] H. Mühleisen and C. Bizer. Web data commons-extracting structured data from two large web corpora. *LDOW*, 937:133–145, 2012.

[13] E. Ostrom. *Governing the commons*. Cambridge University Press, 2015.

[14] P. M. Schwartz. Property, privacy, and personal data. *Harv. L. Rev.*, 117:2056, 2003.

[15] J. Snell and E. Prodromou. Activity streams 2.0. *Working Draft WD-activitystreams-core-20150722*, 2015.

[72] See NEXTLEAP co-ordinator Harry Halpin's talk at the Web 3.0 Summit to the blockchain community: `https://www.youtube.com/watch?v=6PPv3W0InUM`

[16] C. Troncoso, M. Isaakidis, G. Danezis, and H. Halpin. Systematizing decentralization and privacy: Lessons from 15 years of research and deployments. *Proceedings on Privacy Enhancing Technologies*, 2017(4):404–426, 2017.

[17] D. Vrandečić and M. Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.