A Critical Look at Decentralized Personal Data Architectures

Arvind Narayanan
relax@stanford.edu

Solon Barocas
solon@nyu.edu

Vincent Toubiana
vincent.toubiana@alcatel-lucent.com

Helen Nissenbaum
hfn1@nyu.edu

Dan Boneh
dabo@cs.stanford.edu

February 20, 2012

ABSTRACT

While the Internet was conceived as a decentralized network, the most widely used web applications today tend toward centralization. Control increasingly rests with centralized service providers who, as a consequence, have also amassed unprecedented amounts of data about the behaviors and personalities of individuals.

Developers, regulators, and consumer advocates have looked to alternative decentralized architectures as the natural response to threats posed by these centralized services. The result has been a great variety of solutions that include personal data stores (PDS), infomediaries, Vendor Relationship Management (VRM) systems, and federated and distributed social networks. And yet, for all these efforts, decentralized personal data architectures have seen little adoption.

This position paper attempts to account for these failures, challenging the accepted wisdom in the web community on the feasibility and desirability of these approaches. We start with a historical discussion of the development of various categories of decentralized personal data architectures. Then we survey the main ideas to illustrate the common themes among these efforts. We tease apart the design characteristics of these systems from the social values that they (are intended to) promote. We use this understanding to point out numerous drawbacks of the decentralization paradigm, some inherent and others incidental. We end with recommendations for designers of these systems for working towards goals that are achievable, but perhaps more limited in scope and ambition.

1. BRIEF HISTORICAL OVERVIEW

The search for alternatives to centralized aggregation of personal data began in the late 1990s which saw a wave of so-called 'negotiated privacy techniques' including commercial 'infomediaries' [24, 16]. These entities would store consumers' data and help facilitate the drafting of contracts that set the terms of the exchange and use of data. The 1999 book *Net Worth* [23] galvanized both industry and privacy advocates, generating hopes for a future in which privacy problems could be solved through a mix of decentralized storage and private contracts, potentially obviating the need for privacy law or even the adoption of fair information practices [10, 60].

Within five years, nearly all of this excitement had faded and all commercial (Persona, Privada, Lumeria, etc.) and community (P3P) initiatives had floundered [1] — some in truly spectacular fashion, such as AllAdvantage. And yet, by the end of the decade, many new initiatives and projects that shared almost identical goals emerged. Vendor Relationship Management (VRM) [35] has gained steady momentum as a general set of principles that aim simultaneously to improve user privacy, enhance customer autonomy, and increase market efficiency through a combination of mechanisms that aggregate data in a single (per-user) repository under users' control and tools to negotiate agreements that would grant outside organizations access to and use of that data.

Parallel efforts to develop so-called personal data stores (PDS), personal data servers, personal data lockers/vaults, and personal clouds [18] have focused more narrowly on the platforms and protocols to support unified repositories of user data that could be managed locally by the user or outsourced to a trusted third party. The impetus for these projects are varied, ranging from user interest in aggregating one's own data in a single location to better derive benefits from their mixing and matching to more explicit interests in privacy (user control) and commerce (a market place for sharing, including possibilities for cash payments in exchange for data) [13].

The similarities between these and earlier efforts can be quite stark: Mydex's recent white paper, "The Case for Personal Information Empowerment" [38], recapitulates much that was described in a white paper released a full decade earlier by Lumeria, a failed infomediary [30]. To describe this as a simple case of "an idea whose time has come" would be to miss the important lessons that these earlier and recurring failures should offer those who wish to pursue decentralized personal data architectures.

Decentralized social networking has been a largely parallel, sometimes overlapping line of development with similar motivations. We subdivide such social networks into federated (ecosystem of interoperable implementations in the client-server model) and distributed (peer-to-peer). The term distributed social networking is frequently but incorrectly used to describe all decentralized social networks.

While some early thinking in the semantic web community could be classified in this category,[1] for the most part decentralized social networking appears not to have anticipated the success of mainstream commercial, centralized social networks, but rather developed as a response to it. Indeed, prominent members of the web community dismissed social networks until 2007–2008 (for example, [27] and [15]) and academic computer scientists appear to have considered it a passing fad as well — in our survey we see a sharp spike in interest among researchers around this time frame.

A series of well-publicized privacy mishaps by Facebook and Google starting in 2009 that reached its crescendo around the 2010 f8 developer conference stirred up interest among the public and policymakers.[2] Perhaps the most well known project that resulted is Diaspora[3], which was funded in excess of $200,000 via the crowd funding platform kickstarter.com. As of this writing Wikipedia lists about 40 decentralized social networks [58], most of which are federated, whereas the academic literature has focused on distributed social networking for natural reasons, since those present more research challenges.

2. REPRESENTATIVE SURVEY

Rather than attempt an exhaustive survey, in this section we list the key ideas that have been explored in the course of developing decentralized designs. There has been a great fecundity of creative and complex ideas in this space spanning the realms of technology, law and economics; we are unable to present them in detail due to space constraints. We refer the reader to the cited works.

The core idea of an infomediary is that of a trusted third party that interfaces between the user and commercial entities such as marketers [23]. Users' personal data can be manually given to the infomediary, as in Lumeria, or collected through passive monitoring, as in AllAdvantage and other systems [20]. That information can then be utilized without explicit monetization (Mydex, etc.), or users can be paid for their data (AllAdvantage, Bynamite [29], etc). It has variously been argued that telecommunications providers [55, 4], banks [9] and other parties such as providers of home entertainment set-top boxes are ideally suited to play the role of the intermediary. An infomediary might also enable a targeted *attention market* [39] based on user preferences.

Kang et al. introduce the intriguing idea of *licensing* intermediaries to increase their trustworthiness [28]. In the other direction, Vendor Relationship Management systems largely eliminate the infomediary as a separate entity, and instead replace it with a software agent [35]. Some software intermediaries like Adnostic use cryptography to achieve additional privacy properties [54]. Other ideas for improving privacy include fine-grained access control lists [37].

Both VRM and infomediary systems often emphasize benefits to the firm from the intermediated nature of the exchange. Goldman [21] envisions that software agents will make marketing messages perfectly relevant, eliminating externalities from wasted attention. By Coase's theorem [34], this will lead to a socially optimal level of marketing.

Turning to social networks, the key challenge of distributed social networks is hosting and message transfer. One solution is to encrypt messages and store them in a distributed hash table [8, 2]. Another is "social replication": messages are stored in plaintext in a redundant manner by those who have access rights (typically friends of the message poster) [49]. Message passing sometimes exploits the relationship between the social graph and the topology of the physical network [25, 8].

Another frequent goal is keeping edges of the graph secret, for which various solutions have been proposed: a cryptographic approach [5], anonymous routing [14] and friend-to-friend networks such as Freenet in 'darknet' mode [12]. Persona takes the cryptographic heavy-lifting a step further to enable fine-grained access control using attribute-based encryption [6].

Other models for hosting have been explored. In vis-a-vis, each user owns an EC2 virtual host that is active at all times [48], whereas FreedomBox[4] proposes cheap plug computers. Lam et al. have proposed email as a backend [19] and ephemeral networks on smartphones [17]. Unhosted[5] proposes separating data from code, but keeping both in the cloud. Along similar lines, Frenzy[6] is a distributed social network software with Dropbox as the backend. Polaris proposes reducing existing social networks such as Youtube and Twitter to datastores and layering a social network on top, with smartphones providing access control management interfaces [59].

Finally, federated social networks aim to create an ecosystem of standards-based interoperable implementations of social networks. Some designs such as Diaspora are a hybrid between distributed and federated. OStatus, being coordinated by the W3C, represents an interesting approach to standardization for federated microblogging: it references a suite of existing protocols rather than developing them from scratch.

3. CLASSIFICATION

Table 1: The four types of architectures that are the subject of our study

	Commerce, Health etc.	Social Networking
Self-hosted	PDS / VRM	Distributed
Outsourced	Infomediary	Federated

We emphasize that the division in Table 1 is only meant to provide the reader with a rough mental map and is far from precise. The vertical axis, in particular, is closer to a spectrum than a strict division. The terms Personal Data Store and Vendor Relationship Management do not appear to have a single definition. Also, some PDS projects are application-

[1]The Internet Archive lists a version of the *Friend of a Friend* (FOAF) project (www.foaf-project.org) from August 2003, and other efforts may be older.

[2]For an article typifying public opinion during that period, see [45].

[3]`https://joindiaspora.com/`

[4]`http://freedomboxfoundation.org/`

[5]`http://unhosted.org/`

[6]`http://frenzyapp.com/`

agnostic, but these tend to be software libraries/platforms rather than complete user-facing systems.

Towards a finer-grained classification and understanding of different projects, we propose the following (non-independent) axes that are components of what it means for an architecture to be decentralized.

1. Locus of data hosting: this could be remote (centralized), by a trusted third party (infomediary), distributed (peer-to-peer), or local (i.e., on the user's device).
2. Open standards vs. proprietary.
3. Open vs. closed-source implementations.
4. Data portability: Data export (for users), APIs (for third parties), or none.

The above are technical characteristics; one might also try to classify systems in terms of the social values they espouse. We discuss four in particular.

1. Privacy: According to Nissenbaum [41, 40], systems that attempt to preserve privacy should attempt to preserve the integrity of the context in which actors engage with each other. They should do this by ensuring that information flows respect the norms of the context. To the degree that systems better model and mediate appropriate information flows, they will advance the privacy interests of their users. This view will inform the discussion in Section 4.1.
2. Utility: We refer to the overall social benefit of the system, in the sense of welfare maximiation in economics. One way to achieve increased utility is through greater interoperability or data portability.
3. Cost: Cost encompasses hosting and bandwidth costs as well as software development and maintenance costs. Centralized and decentralized systems behave very differently: in the former case there is typically a single entity that bears all the costs whereas in the decentralized setting it can be split among users and various software creators and service providers. Comparing these alternatives may therefore be tricky.
4. Innovation: We must also consider how quickly different systems are able to evolve and adapt. Some have argued that open standards catalyze innovation while others point to the time and monetary costs of standardization. The strength of the business model, the extent of market competition, the ability to harness and analyze data, and legal compliance requirements are some of the other factors that affect how conducive a system is to innovation.

Values may not be immediately deducible from the technical design of a system, but may instead only be observable empirically. Indeed, we suggest that much of the reason for what we see as overenthusiastic claims about decentralized systems is that design characteristics have been confused with values. We discuss two prominent cases in detail in Sections 4.1 and 4.2. Moreover, we doubt whether any architecture could optimize for all values simultaneously.

4. DRAWBACKS OF DECENTRALIZATION

In this section we present some underappreciated drawbacks of decentralized architectures. Not all of these apply to all types of systems, nor is any of them individually a decisive factor. But collectively they may help explain why decentralization faces a steep road ahead, and why even if adopted, decentralization will not necessarily provide all the benefits that its proponents believe will automatically flow from it.

An architecture without a single point of data aggregation, management and control has several **technical** disadvantages. First is functionality: there are several types of computations that are hard or impossible without a unified view of the data. Detection of fraud and spam, search, collaborative filtering, identification of trending topics and other types of analytics are all examples. Decentralized systems also suffer from inherently higher network unreliability, resulting in a tradeoff between consistency and availability (formalized as the CAP theorem [57]); they may also be slower from the user's point of view.[7] The need for synchronized clocks and minimizing data duplication are other challenges.

The benefits and costs of standardization are a prominent socio-technical factor. Many decentralized systems depend on multiple interoperating pieces of software, which requires standardization of technical protocols, design decisions, etc. On the one hand, such an ecosystem could promote long-term innovation; on the other hand, these processes (e.g., HTML5) move at a far slower pace than Facebook or an ad network which can roll out features over the timespan of days or weeks. Shapiro notes two benefits of standardization: greater realization of network effects and protection of buyers from stranding, and one cost: constraints on variety and innovation, and argues that the impact on competition can be either a benefit or a cost [50].

Let us now turn to **economics**. Centralized systems have significant *economies of scale* which encompasses hosting costs, development costs and maintenance costs (e.g., combating malware and spam),[8] branding and advertising [42]. A related point in the context of social networks: we hypothesize that the network effect is stronger for centralized systems due to tighter integration.

Path dependence is another key economic issue: even if we assume that centralized and decentralized architectures represent equally viable equilibria, which one is actually reached might be entirely a consequence of historical accident. Most of the systems under our purview – unlike, say, email – were initially envisioned as commercial applications operating under central control, and it is unsurprising they have stayed that way.

The theory of *unraveling* suggests that infomediaries in particular might *not* in fact represent a stable equilibrium. For an infomediary to succeed, consumers and businesses must

[7]Google reports that users exposed to an additional delay of as little as 100ms performed a statistically significantly smaller number of searches [44].

[8]Facebook has built a highly sophisticated real time "immune system" which relies in part on human operators [51].

transact through the intermediary rather than directly with each other. But either side of this market might see participants iteratively defecting, resulting in unraveling of the market. Chen et al. discuss how this might happen from the businesses' side [11], and Peppet discusses it from the consumer side [43]. However, it is not fully clear why many types of intermediaries have taken hold in many other markets — employment agents, goods appraisers, etc — but not in the market for personal data.

A variety of **cognitive** factors hinder adoption of decentralized systems as well. First, the fact that decentralized systems typically require software installation is a significant barrier. Second, more control over personal data almost inevitably translates to more decisions, which leads to cognitive overload. Third, since users lack expertise in software configuration, security vulnerabilities may result. A related point is that users may be unable to meaningfully verify privacy guarantees provided through cryptography.

Finally, we find that decentralized social networking systems in particular fare poorly in terms of mapping the norms of information flow. Access control provides a very limited conception of privacy. We provide several examples. First is the idea of "degrees of publicness." For example, on Facebook a post may be publicly visible, yet the site has defenses to stop crawlers which prevents the post ending up in a search engine cache, so that the user may meaningfully hide or delete the post later if they so choose. Second, in current social networks privacy is achieved not only through technical defenses but also through "nudges" [36]. When there are multiple software implementations, users cannot rely on their friends' software providing these nudges. Third, distributed social networks reveal users' content consumption to their peers who host the content[9] (unless they have a "push" architecture where users always download accessible content, whether they view it or not, which is highly inefficient.) Finally, decentralized social networks make reputation management and "privacy through obscurity" (in the sense of [26]) harder, due to factors such as the difficulty of preventing public, federated data from showing up in search results.

4.1 On Control over Personal Data

We now discuss two drawbacks in detail to illustrate the difference between architectural decisions and social values that they are often implicitly assumed to promote. The first is the distinction between control over hosting and privacy. To elucidate this we present a thought experiment.

What does it mean for users to truly host and control their personal data, while still being able to participate in activities such as social networking and personalized commerce? Compared to using Facebook, hosting one's data on a personal EC2 instance certainly puts the user in greater control, but Amazon will turn over user data in response to a subpoena or court order [3].

For any hope of absolute control, users must, at a minimum, host data on their own device resident on their physical property. This is already considerably at odds with the reality of today's consumer Internet: bandwidth to the home is often asymmetric, or connectivity is restricted in other ways (NATs, firewalls), and few individuals possess always-on devices capable of running web services.[10]

Furthermore, the software running the services must be open-source, and be audited by third-party certification authorities, or by "the crowd". Silent auto-updates, which is the model that client-side software is increasingly moving towards, would be difficult due to the auditing requirement, perhaps prohibitively so.

Further still, *hardware* might have backdoors, and therefore needs an independent trust mechanism as well. The user also needs the time and knowhow to configure redundant backups, manage software security, etc. Finally almost all decentralized architectures face the the problem of "downstream abuse" which is that the user has no technical means to exercise control over use and retransmission of data once it has been shared [47].

This thought experiment shows that absolute control is impossible in practice. Further, it suggests that control over information is probably not the right conceptualization of privacy, if privacy is the end goal.

4.2 Open standards and Interoperability

Interoperability is a laudable goal; it could enhance social utility, as we have mentioned earlier. However, it has frequently been reduced to the notion of open standards. We argue here that while open standards are a prerequisite for interoperability, there is a big gap between the two. In particular, the efforts at federated social networking all follow open standards, but their actual interoperability status in practice appears to be poor [56]. Let us examine why this is the case.

One major impediment is that there are too many standards to choose from. For the most basic, foundational component — identity — there are many choices: OpenID, WebID and others. While it is possible to connect these to each other, it requires extra effort. As we get to more complex (but still basic) functionality such as federation of messages, we find on the one hand Atom/PubSubHubbub etc. and the OStatus suite[11] on top of it, and on the other hand XMPP and the Wave federation protocol[12] on top of it. It appears that the former is gradually winning out, but this is a slow process.

The second major impediment is that as soon as we get past the basics like identity, friendship and status updates, there is an incredible array of parameters to nail down. Take the apparently trivial issue of what formatting is allowed in a status update. Unless two systems agree on the same standard, they are not interoperable because users of one will see malformatted messages originating from the other. Needless to say, centralized platforms have a large and ever-

[9] This is a particularly serious problem for systems like Contrail [52].

[10] It remains to be seen if smartphones will become practical for this use-case.

[11] `http://ostatus.org/`

[12] `http://www.waveprotocol.org/`

increasing set of features — photos, video chat, polls, to name a few — all of which would require standardization in the federated context. Finally, access control in a federated setting presents hard technical challenges.

The practical upshot is that the only suite of standards that shows any signs of meaningful interoperability is StatusNet[13] — microblogging is both text based, largely eliminating the formatting issue, and typically public, sidestepping the access-control issue — although identi.ca remains the only implementation with meaningful adoption. Even though this system limits status updates to text, a version of the formatting problem still plagues it: identi.ca restricts updates to 140 characters in an attempt to maintain some interoperability with Twitter!

We conclude that while federated social networks have the potential to converge on a reasonably interoperable collection of software — subject to the caveats of differing feature sets and parameters — it is not simply a matter of making some technical decisions, but instead needs serious developer commitment as well as the involvement of standards bodies with significant authority.

5. RECOMMENDATIONS

Based on our analysis above, we offer the following recommendations for developers of decentralized systems.

1. Consider the economic feasibility of your design. In particular, are there entities with the economic incentive to play the various roles that are called for? This has perhaps been the most common reason for the lack of adoption of past proposals and projects.

2. Pay heed to conceptual fidelity. Are you shooting at the right target? Do people have the values you think they do? Do they really want the features/benefits you claim they want? As one example, there have been a multiple of projects that attempt encrypted communication over Facebook and other social networks (NOYB [22], FlyByNight [32], Lockr [53], FaceCloak [33], Scramble! [7], etc.), but the lack of adoption suggests that the usability costs do not outweigh the benefits to users.

3. Incorporate other notions of regulability [61, 31]. Many decentralized systems represent an extreme choice: they seek to achieve privacy and other properties purely through technology, ignoring socio-legal approaches. This extreme may not be optimal.

4. Offer advantages other than privacy to users. Privacy is always a secondary feature — while it might tip the balance between two competing products, users rarely pick a product based on privacy alone. For example, distributed social networking can enable some location-specific functionalities through peer-to-peer networking even when there is no Internet access.

5. Design with standardization in mind. One of the disadvantages we have identified is the proliferation of non-interoperable systems. Open standards are not enough: developers must actively prioritize interoperability and write and maintain glue code to interface with other systems.

6. Target limited feature sets. A system like Facebook is a large, complex moving target. Attempting to create a decentralized version of it is a futile endeavor. Instead, systems that embody the 'minimum viable product' strategy might succeed better in the market. Decentralized microblogging appears to be a relatively attainable goal at the present time, and censorship resistance is a goal for which there is much demand.

7. Work with regulators. As numerous law/economics scholars have pointed out, market solutions appear to underprovide privacy and regulation can help tweak the environment to address this imbalance [46]. Those who wish to see the personal data ecosystem flourish would do well to support regulatory interventions such as transparency and opt-out that can help level the playing field between centralized and decentralized systems.

[13] `http://status.net/`

6. CONCLUSION

In this position paper we have taken a look back at the efforts to build decentralized personal data architectures motivated either by discontent with the status quo, or as a better way to organize information markets and leverage new commercial opportunities, or a combination of both. We hope we have provided some mental clarity to the reader on the similarities, differences and common themes between the various systems and brought fresh perspective to the question of why they have largely floundered.

We hope to kick off a more tempered discussion of the future of personal data architectures in both scholarly and hobbyist/entrepreneurial circles, one that is informed by the lessons of history. There is much work to be done along these lines — application of economic theory can shed light on questions such as the relative strength of network effects in centralized vs. decentralized systems. Empirical methodology such as user and developer interviews would also be tremendously valuable. While we have provided some suggestions for developers, in the future we hope to identify specific application domains that are relatively amenable to the adoption of decentralized architectures, as well as to provide concrete recommendations for policymakers who might wish to foster a different market equilibrium.

Acknowledgement. The first author would like to thank Monica Lam and the other members of the MobiSocial project for enlightening discussions, Deirdre Mulligan, Nick Doty and Jennifer King for helping develop ideas on a multi-factor approach to privacy, Alejandro Molnar for general education about economics, Alessandro Acquisti for sharing his bibliography on the economics of privacy and numerous online commenters for perspectives, links and information.

7. REFERENCES

[1] M. S. Ackerman. The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Hum.-Comput. Interact.*, 15(2):179–203, Sept. 2000.

[2] L. M. Aiello and G. Ruffo. Lotusnet: tunable privacy for distributed online social network services. *Computer Communications*, In Press,, 2010.

[3] Amazon, Inc. AWS Customer Agreement. `http://aws.amazon.com/agreement/`.

[4] I. Ayres and M. Funk. Marketing privacy. *The Yale Journal on Regulation*, 20(1):77–137, January 2003.

[5] M. Backes, M. Maffei, and K. Pecina. A security api for distributed social networks. In *NDSS*, 2011.

[6] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin. Persona: An online social network with user-defined privacy. *Computer*, 39(4):135–146, 2009.

[7] F. Beato, M. Kohlweiss, and K. Wouters. Scramble! Your Social Network Data. In *PETS*, pages 211–225, 2011.

[8] S. Buchegger, D. Schioberg, L. H. Vu, and A. Datta. PeerSoN: P2P Social Networking - Early Experiences and Insights. In *Proceedings of the Second ACM Workshop on Social Network Systems Social Network Systems 2009, co-located with Eurosys 2009*, Nurnberg, Germany, March 31 2009.

[9] Carol Matlack. Who Do You Trust More with Your Data: Facebook or a Bank? BusinessWeek, 2012.

[10] J. Catlett. Panel on infomediaries and negotiated privacy techniques. In *Proceedings of the tenth conference on Computers, freedom and privacy: challenging the assumptions*, CFP '00, pages 155–156, New York, NY, USA, 2000. ACM.

[11] Y. Chen, G. Iyer, and P. V. Padmanabhan. Referral Infomediaries and Retail Competition. *Review of Marketing Science Working Papers*, 1(2), November 2001.

[12] I. Clarke, S. G. Miller, T. W. Hong, O. Sandberg, and B. Wiley. Protecting free with freenet. *Internet Computing IEEE*, 6(February):40–49, 2002.

[13] P. D. E. Consortium. The Startup Circle. `http://personaldataecosystem.org/2011/06/startup/`, 2011.

[14] L. Cutillo, R. Molva, and T. Strufe. Safebook: A privacy-preserving online social network leveraging on real-life trust. *IEEE Communications Magazine*, 47(12):94–101, 2009.

[15] Dave Winer. Why Facebook Sucks. `http://scripting.com/stories/2007/10/13/whyFacebookSucks.html`, 2007.

[16] A. Dix. Infomediaries and negotiated privacy techniques. In *Proceedings of the tenth conference on Computers, freedom and privacy: challenging the assumptions*, CFP '00, pages 167–, New York, NY, USA, 2000. ACM.

[17] B. Dodson and M. Lam. Musubi: A mobile privacy-honoring social network. 2010.

[18] E. Drummond. The Personal Cloud. `http://equalsdrummond.name/2011/02/07/the-personal-cloud/`, 2011.

[19] M. H. Fischer and M. S. Lam. Email Clients as Decentralized Social Apps in Mr . Privacy. In *Proceedings of the fourth Hot Topics in Privacy Enhancing Technologies , co-located with Eurosys 2009*, July 2011.

[20] B. Givens. Infomediaries and negotiated privacy: resources. In *Proceedings of the tenth conference on Computers, freedom and privacy: challenging the assumptions*, CFP '00, pages 165–166, New York, NY, USA, 2000. ACM.

[21] E. Goldman. A coasean analysis of marketing. *Wisconsin Law Review*, (4):1151–1221, 2006.

[22] S. Guha, K. Tang, and P. Francis. Noyb: privacy in online social networks. In *Proceedings of the first workshop on Online social networks*, WOSN '08, pages 49–54, New York, NY, USA, 2008. ACM.

[23] J. Hagel and M. Singer. *Net worth: shaping markets when customers make the rules*. Harvard Business School Press, 1999.

[24] J. Hagel, III and J. F. Rayport. The coming battle for customer information. In *Creating value in the network economy*, pages 159–171. Harvard Business School Press, Boston, MA, USA, 1999.

[25] L. Han, B. Nath, L. Iftode, and S. Muthukrishnan. Social butterfly: Social caches for distributed social networks. In *SocialCom/PASSAT*, pages 81–86, 2011.

[26] W. Hartzog and F. D. Stutzman. The Case for Online Obscurity. *Privacy Papers for Policy Makers*, 2011.

[27] Jeff Atwood. Avoiding Walled Gardens on the Internet. `http://www.codinghorror.com/blog/2007/06/avoiding-walled-gardens-on-the-internet.html`, 2007.

[28] J. Kang, K. Shilton, D. Estrin, J. Burke, and M. Hansen. Self-surveillance privacy. *Iowa Law Review*, 97:809, December 2010.

[29] Kashmir Hill. Names You Need To Know in 2011: Bynamite. Forbes, 2011.

[30] F. Labalme and J. Duwaik. An infomediary approach to the privacy problem. *Feb*, 9:24, 1999.

[31] L. Lessig. *Code and other laws of cyberspace*. Basic Books, 1999.

[32] M. M. Lucas and N. Borisov. FlyByNight: mitigating the privacy risks of social networking. In *Proceedings of the 7th ACM workshop on Privacy in the electronic society*, WPES '08, pages 1–8, New York, NY, USA, 2008. ACM.

[33] W. Luo, Q. Xie, and U. Hengartner. FaceCloak: An Architecture for User Privacy on Social Networking Sites. In *Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 03*, pages 26–33, Washington, DC, USA, 2009. IEEE Computer Society.

[34] A. Marciano. Ronald Coase, 'The Problem of Social Cost' and The Coase Theorem: An anniversary celebration. *European Journal of Law and Economics*, 31(1):1–9, 2010.

[35] A. Mitchell, I. Henderson, and D. Searls. Reinventing direct marketing — with vrm inside. *Journal of Direct Data and Digital Marketing Practice*, 10(1):3–15, 2008.

[36] D. Mori. Privacy nudges protect information. `http://thetartan.org/2010/3/22/scitech/privacynudges`,

2010.

[37] M. Mun, S. Hao, N. Mishra, K. Shilton, J. Burke, D. Estrin, M. Hansen, and R. Govindan. Personal Data Vaults: a locus of control for personal data streams. 2010.

[38] Mydex. The case for personal information empowerment : The rise of the personal data store, 2010.

[39] NASDAQ/Edgar Online. AllAdvantage IPO filing. http://ipo.nasdaq.com/TextSection.asp?cikid=71762&fnid=3262&sec=bd, 1999.

[40] H. Nissenbaum. *Privacy in Context - Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2010.

[41] H. F. Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79(1):119–157, 2004.

[42] Y. Peles. Economies of scale in advertising beer and cigarettes. *The Journal of Business*, 44(1):32–37, 1971.

[43] S. R. Peppet. Unraveling Privacy: The Personal Prospectus and the Threat of a Full Disclosure Future. *Northwestern University Law Review*, 105(3), 2011.

[44] G. Research. Speed matters. http://googleresearch.blogspot.com/2009/06/speed-matters.html, 2009.

[45] Ryan Singel. Facebook's Gone Rogue; It's Time for an Open Alternative. http://www.wired.com/epicenter/2010/05/facebook-rogue/, 2010.

[46] P. M. Schwartz. Privacy and democracy in cyberspace. *Vanderbilt Law Review*, 52:1609–1701, 1999.

[47] P. M. Schwartz. Property, Privacy, and Personal Data. *Harvard Law Review*, 117(7):2055–2128, 2004.

[48] A. Shakimov, H. Lim, K. Li, D. Liu, and A. Varshavsky. *Vis-a-Vis: Privacy-Preserving Online Social Networking via Virtual Individual Servers*. IEEE, 2011.

[49] A. Shakimov, A. Varshavsky, L. P. Cox, and R. Cáceres. Privacy, cost, and availability tradeoffs in decentralized osns. In *Proceedings of the 2nd ACM workshop on Online social networks*, WOSN '09, pages 13–18, New York, NY, USA, 2009. ACM.

[50] C. Shapiro. Setting Compatibility Standards: Cooperation or Collusion? In R. Dreyfuss, D. Zimmerman, and H. First, editors, *Expanding the boundaries of intellectual property: innovation policy for the knowledge society*. Oxford University Press, 2001.

[51] T. Stein, E. Chen, and K. Mangla. Facebook immune system. In *Proceedings of the 4th Workshop on Social Network Systems*, SNS '11, pages 8:1–8:8, New York, NY, USA, 2011. ACM.

[52] P. Stuedi, I. Mohomed, M. Balakrishnan, Z. M. Mao, V. Ramasubramanian, D. Terry, and T. Wobber. Contrail: Enabling decentralized social networks on smartphones. In *Middleware*, pages 41–60, 2011.

[53] A. Tootoonchian, S. Saroiu, Y. Ganjali, and A. Wolman. Lockr: better privacy for social networks. In *CoNEXT*, pages 169–180, 2009.

[54] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas. Adnostic: Privacy preserving targeted advertising. In *In NDSS*, 2010.

[55] Vodafone. Rethinking Personal Data. http://www.vodafone.com/content/dam/vodafone/about/privacy/vodafone_rethinking_personaldata.pdf, 2011.

[56] W3C Federated Social Web Incubator Group. Social Web Acid Test - Level 0. http://www.w3.org/2005/Incubator/federatedsocialweb/wiki/SWAT0, May 2011.

[57] Wikipedia. CAP theorem. http://en.wikipedia.org/w/index.php?title=CAP_theorem&oldid=472196147. [Accessed 4-February-2012].

[58] Wikipedia. Distributed social network. [Accessed 4-February-2012]. http://en.wikipedia.org/w/index.php?title=Distributed_social_network&oldid=471838360.

[59] C. Wilson, T. Steinbauer, G. Wang, A. Sala, H. Zheng, and B. Y. Zhao. Privacy, availability and economics in the polaris mobile social network. In *Proc. of HotMobile*, Phoenix, AZ, March 2011.

[60] Wired. The Dawn of the Infomediary. http://www.wired.com/techbiz/media/news/1999/02/18094, 1999.

[61] J. Zittrain. *The Future of the Internet–And How to Stop It*. Yale University Press, New Haven, CT, USA, 2008.