Beyond Walled Gardens: Open Standards for the Social Web

Harry Halpin

Institute for Communicating and Collaborative Systems
University of Edinburgh
2 Buccleuch Place
Edinburgh, United Kingdom
H.Halpin@ed.ac.uk

Now that the Social Web has finally reached truly widespread adoption, the question remains: Why can't users have their data back? To flip the question, what could researchers discover and application-developers create if they had access to the masses of social data currently spread throughout the Web? Luckily, the building-blocks that allow us to open up these closed walls of data exist right now: all that is missing is a strategy for putting it together. As has been said before: You may not be interested in strategy, but strategy is interested in you.

One weak point is conceptual: Social data portability and privacy are usually viewed as opposing forces. Yet data portability and privacy are mutual benefits that a framework for a mature Social Web could bring users. Today, the walled garden of data fundamentally leads to less security and privacy for users. For example, the lack of portability does not imply a lack of privacy: the data of users may be data-mined and is all-too- portable, for it can be easily sold to parties unknown to the users without their explicit knowledge. Furthermore, the lack of portability has made it common practice for many social web services to ask users to give third-party services access e-mail inboxes, an insecure practice that has already easily led to identity theft. By forcing users to have their data spread throughout the Web under multiple accounts, users often just repeat passwords and user-names, leading to insecure transactions. Solving these real-world problems should not be rocket science, although it may be Web science.

Technically, solutions to all these problems already exist. SAML (Security Assertion Markup Language) in general, and OpenID, provide a usable framework for multi-site log-ins. OAuth can provide authenticated API access. For data portability, a host of incompatible APIs and data formats exist, ranging from OpenSocial to the Contact API, from the XFN microformat to FOAF. One large question that must be answered is how can all these different standards be harmonized, and on what level of abstraction? Given the large amount of work already put into these technologies, instead of asking for a single API to be adopted, a more sensible strategy would be move to an extensible and simple modeling framework, and rather shockingly the Semantic Web may very well be the best solution out there.

Yet, the Semantic Web has its own host of problems, and despite the years of research, very little research has gone into maturing the social side of the Semantic Web. This points out a fundamental flaw in the design of RDF as it

stands today: While publishing triples in the wild may do for publicly-available Linked Data, this model of deployment will not work for Social Web. Instead, somehow privacy and data provenance must be built into the very core of the data-format, and be easily accessible. Furthermore, the work on trust needs to go beyond Goldbeck's famous Trust Ontology. How does the Semantic Web and identity providers interact? How can we support both identity integration and multiple profiles for different uses? And on a very basic level, what are the mappings between vCard, XFN, and FOAF? What precisely is the core of social data that should be standardized, and what other components should be left to develop in a decentralized manner? These questions are seemingly simple, and one is unlikely to get a dissertation working on them. Without concrete answers and running code to solve these practical questions, the Semantic Web vision is unlikely to take off.

Luckily, the World Wide Web Consortium provides just such a process where academia, industry, and developers can discuss the future of the Social Web, create a strategy, and then implement it. Furthermore, in a consensus-driven manner, various parties can get on board, without the fear of patent trolls due to the W3C's Royalty-Free Patent Policy. Lastly, the W3C process can help guarantee that various other parts of the Web, like the Mobile Web, can stay involved. What is needed from academia is that the research priorities of the Semantic Web move beyond its roots in classical artificial intelligence to the problem of creating a framework for collective intelligence on the Social Web.