

Beyond Privacy

Andreas Harth

Karlsruhe Institute of Technology

April 15, 2011

With the proliferation of computers, an ever increasing amount of data is being stored and processed. The Internet and the Web have enabled the easy publication of and access to data. Semantic Web [4] technologies – in particular Linked Data [6] and ontologies [9] – promise an even easier path to accessing and integrating data to make our economies more efficient, use less natural resources and ultimately attain more knowledge about the world.

These technological advances are often touted as a goal worth attaining, while little thought is given to how the implementation of these technologies might impact everybody, from individuals to companies to governments. A brave new world, where information that wants to be free is actually liberated, may actually hold changes in store which not everybody currently anticipates in their total breadth and depth. This position paper aims at giving an account on current developments and providing food for thought as to what impact the new data integration technologies might have. While the developments have already happened and will likely continue, one ought to anticipate the consequences and one should, depending on one's character, either brace for or embrace what is coming.

A significant portion of the data that is being made available comprises personal data, which accumulates in both open and closed environments. For example, web users often supply personal data when signing up for services on web sites which are being made publicly available and are actively used for commercial purposes. Websites which require users to log in to access content, too, are target for data collection for commercial purposes [2]. Companies actively use data recorded as users interact with web sites (“data exhaust”) to generate value [1]. Even when data is stored in data management systems residing behind firewalls, the mere fact that more personal data is available (e.g., recorded via mobile phones [5]) may lead to increased use of the data for profit and heightened potential for abuse, as data may be applied beyond the intended and agreed-upon or expected purpose. For example, rouge employees might view records out of personal interest or leak data. In addition, when and how official organisations access and use personal data is often opaque. Once data is being recorded digitally, the genie is out of the bottle, and it's unclear for the individuals concerned who is using their data. Questions around data ownership also arise when companies develop new ideas, products and services based on the insights mined from personal data [7].

Similar developments can be witnessed in the commercial environment. Competitive Intelligence aims at attaining information about competitors via various means, one of them being the use of information technology. While companies try their best to obfuscate information they give to the public (for example in reports to the Securities and Exchange commission), collating data from a large and diverse number of sources may give competitors an increasingly clear picture on the inner workings of the company. The corollary is that companies actively leveraging data at their disposal may gain an advantage over their competition.

Privacy and transparency are different sides of the same coin: somebody's transparency affects somebody else's privacy. Greater availability of data, both on the open Web and behind firewalls, inevitably leads to dormant potential for abuse, but at the same time offers great potential for innovation. Data management and integration systems are incredibly useful and make lives easier, the economy more productive and scientific discoveries possible. Given the new data technologies, we have to rethink traditional notions of privacy [10]. Researchers have to be overt about how newly developed systems may impinge upon the rights of data owners; we need to actively confront issues related to privacy [3] to avoid realising the consequences of our actions when it's too late [8].

References

- [1] Data, data everywhere - A special report on managing information. *The Economist*, Feb 25 2010. <http://www.economist.com/node/15557443>.
- [2] J. Angwin and S. Stecklow. 'Scrapers' Dig Deep for Data on the Web. *The Wall Street Journal*, October 12 2010. <http://online.wsj.com/article/SB10001424052748703358504575544381288117888.html>.
- [3] T. Berners-Lee. Long Live the Web: A Call for Continued Open Standards and Neutrality. *Scientific American*, Nov. 2010. <http://www.scientificamerican.com/article.cfm?id=long-live-the-web>.
- [4] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web: *Scientific American*. *Scientific American*, May 2001.
- [5] K. Biermann. Betrayed by our own data. *Zeit Online*, Mar 26 2011. <http://www.zeit.de/datenschutz/malte-spitz-data-retention>.
- [6] C. Bizer. The web of linked data: a global public dataspace on the web: Webdb 2010 keynote. In *Proceedings of the 13th International Workshop on the Web and Databases*, pages 1–1, 2010.
- [7] B. Kreit. The Future of the Value of Data, Mar 2011. <http://www.iftf.org/node/3804>.
- [8] W. L. Laurence. Now we are all sons-of-bitches. *Science News*, 98(2):39–41, 1970. <http://www.jstor.org/stable/3955694>.
- [9] S. Staab and R. Studer, editors. *Handbook on Ontologies*. International Handbooks on Information Systems. Springer, 2004.
- [10] D. J. Weitzner. The transparency paradox: Privacy design strategies for open information networks. In *Proceedings of Location Privacy Workshop: Individual Autonomy as a Driver of Design*, August 2004. <http://www.w3.org/2004/05/loc-priv-transparency-extab.html>, accessed 2011-03-31.