

Viewpoint

Privacy on the Data Web

Considering the nebulous question of ownership in the virtual realm.

THE WORLD WIDE Web in its current form, linking documents with hyperlinks in an associative network, has led to a number of concerns about issues related to privacy, copyright, and intellectual property.⁵ But the movement away from the linking of *documents* to the linking of *data*, a much more powerful paradigm allowing automation of a greater number of information processing tasks, will test legal and technical regimes still further.

The linked data Web, in which heterogeneous data is brought together from distributed sources relatively seamlessly with user-provided ontologies, allows information about individuals or organizations to be queried despite being collected at different times for different purposes, with different provenances and different formats. The benefits of such a Web are manifest^{5,8} but threats to personal privacy will also increase as boundaries blur between personal information published intentionally, that published conditionally (for example, to specific social networking sites for a specific audience) and information over which the subject has no control.

One way of expressing the dilemma that will face us is to ask the question “who owns all this data?” When it is personal data, surely we do?

Perhaps surprisingly, the answer is no. Even if you enter the data yourself, for example onto some Internet service, you do not own it—the service

generally does. You will have signed up for something in the small print—that is, you will tacitly have consented to handing over the data. Given the highly interactive nature of the Web where one creates data consciously and unconsciously all the time, this consent model will be increasingly stretched over the next few years.

It has always been somewhat flawed, with few limits to the uses to which data is put when consent to process

has been lawfully obtained (and privacy policies may change after one has consented³). Naïve users and minors often treat policies, or terms and conditions, as a tedious box necessary to check to get onto a site, rather than as signing away their rights.⁷ But even when there are no problems of asymmetric information or proportionality, there are social issues to be considered—privacy is not a private matter. It impacts on a series of wider communities.

Art in Development

A social network includes lots of information not directly about you. The information is implicit, but network analysis makes it explicit. The evidence of a network is circumstantial, but an important basis for profiling. For example, if you have a high percentage of gay friends does that mean you are gay? Many people—gay or straight—would find that inference embarrassing.

We do not own our networks. In January 2008, blogger Robert Scoble automatically harvested the names and email addresses of his several thousand Facebook friends, and exported them to another account. The row was resolved amicably in the end—but the outcome was that Scoble’s network was not his to harvest.

Given the benefits of wide access to data, it is appropriate to ask whether “ownership” is the concept needed. In the first place, legal frameworks that define a type of data ownership for the subject are absent—these are facts *about* a person, not copyright material, intellectual property, or trade secrets.

Second, the most important power of ownership is denial of access: if I own something, I can stop you using it. But this undermines the potential of the Web of linked data. In the old days of paper and practical obscurity, the value of information was in its scarcity, but on the Data Web value comes from abundance, the ability to place information in new and unexpected contexts, facilitating what Tim Berners-Lee calls “serendipitous reuse.”⁸ Ensuring data is correct is more valuable than preventing its use. We should also not ignore the opposite pull from rights of access to information, as a corollary to rights of freedom of expression,⁶ while many people and organizations have legitimate interests in access to data.

This is the rationale for data protection, whose aim is not exclusively to protect individuals’ privacy, but rather to balance privacy with the maintenance of the free flow of information, as well as other desirable things for individuals like quality and accessibility.⁹ Under a data protection regime, individuals have the right to inspect and correct information being held about them, in theory allowing them to address issues of incorrectness, inappropriateness, excessiveness, and so on.

It also has the effect of bringing rules

into the area directly—data protection provides controls administered by a regulatory body over how data should be handled. On the other hand, one’s privacy can only be addressed under an ownership regime in court *after* a tort or legal injury had occurred as a result of misuse.

In Europe, the 1981 Council of Europe Convention on data protection was required to reconcile the right of privacy in Article 8 of the European Convention on Human Rights with the right of freedom of expression given in Article 10. The Convention led directly to the EU’s directive on data protection in 1995 (95/46/EC), and to national legislation such as the U.K.’s Data Protection Act of 1998.⁹ Most industrialized nations have some sort of data protection legislation in place, although European laws are probably the most comprehensive.

There are differences between jurisdictions, of which some extend protection to legal entities like companies, others include non-digital information under the remit, others have restricted data protection to public sector data, while still more have argued that information affecting national sovereignty or sociocultural interests should also fall under the banner, with states having rights as well as individuals.

This variation is often cultural; some nations value privacy more than others, Continental Europeans worry about corporations’ access to data, while Anglophone nations tend to be more suspicious of governments, and so on. Yet it also matters economically—some senior business people suspect that such is the value of data that businesses in those states with strong data protection laws, such as Germany, could well lose out to those in jurisdictions with less protection, such as the U.K.

Given the benefits of wide access to data, it is appropriate to ask whether “ownership” is the concept needed.

Different regimes offer different levels of protection. Consider for instance the definition of personal data. Belgium has incorporated the wording of Directive 95/46/EC directly into law, covering anyone who can be identified directly or indirectly from the data, while the U.K. has altered the wording to cover only those who could be identified *by the data controller* from the data. Data that can be used to identify one (such as an IP address) can be collected without data protection legislation in the U.K. as long as the controller has no way of going from IP address to an individual.⁷

Nevertheless, the Web is an opaque place, especially to non-expert users. Putting the onus on the data *subject* to ask for details of how personal data is being used ensures that much will be missed—how many know the right questions to ask about cookies, ISPs, search engines, or browsers? Will it pay regulators to take a stronger stance?

Regulation of the Web is a complex matter, crossing jurisdictions and posing problems for the W3C’s consensus-based standards approach. Regulation generally leverages normality, and is premised on common behavior and shared interpretations of a situation.^{3,10} It is more effective if it goes with the grain of a society’s norms, but online there is no “normal” behavior, as work on the scale-free aspects of the Web has repeatedly demonstrated (recently in Meiss et al.⁴), while user understanding of online situations is highly heterogeneous.

The Web moves so quickly that regulation is risky. It takes time and coordination across borders; by the time rules are in place, behavioral patterns may likely have changed, and all that is left is unintended consequences.⁵ Directive 95/46/EC dates back to 1995, with key updates to cover traffic and location data introduced in 2002. The scale and speed of the Web’s evolution means that carefully considered regulation is rarely timely; the whole privacy-threatening phenomenon of Web 2.0 has arisen since those directives. For example, in social network sites friends sometimes take information that a user had originally characterized as private to them and republish it to their immediate friends. The discipline of Web Science covered recently

in these pages² is an attempt to harness transdisciplinary endeavor to try to understand the complex feedback cycles between the Web and society.

If ownership and regulation are problematic, what to do? We have two proposals, one modest, one a little deeper.

As things stand, privacy is a game for the rich and well informed, creating a digital divide to which one response is to redress the balance by exploring ways in which people can perceive advantage from protecting their privacy. In particular, if we can shift the emphasis from concealment to transparency—from the *concealment* of data from potential users, to *transparency* of how data is being used—we can begin to provide answers to questions like “who is looking at you?” and “what is being said about you?” Data will continue to be gathered, aggregated and graphed, but its use should be clear and traceable. We are of course gesturing toward the work of Daniel Weitzner and colleagues on information accountability, reported in this magazine.¹⁰

With a proper infrastructure in place, it should be possible to construct legal/technical/economic models where people can be recompensed for the use of their data—you could be paid for your clickthroughs. Or perhaps you would require a donation to a cause of your choice in return for your clickthroughs. If others are making money from observation of your activity, it doesn't seem outrageous that you or your nominees should have a slice of the action.

It may be that the commercial thirst for consumer data is about to wane as the global financial crisis undermines advertising, and therefore the business models of many Internet companies. But this idea is just one instance of a more general principle of reciprocity between technology developers and information subjects. If a technology makes public service more efficient, or a business process more profitable, then it should also be used reciprocally to aid the citizen or consumer.

If government officials have better access to data as a result of intrusive technology, then citizens should too—improved data for government implying more freedom of information. A consumer should be able to get

Perhaps we should be talking of the responsibilities of privacy too.

improvements in data protection, for example by being able to use technology to enforce access to information in the many jurisdictions where such enforcement is currently problematic.⁶

As our rights as citizens and as consumers seem to be coming together, markets could be redefined to change the incentives to protect one's own privacy and respect that of others, for example, as with principles such as ‘the polluter pays.’ The analogy with pollution is suggestive for our more fundamental idea—an invasion of privacy has things in common with pollution, in particular that the individual benefits and costs do not capture the full social costs.

In many jurisdictions, particularly common law ones, the complexities of privacy are dealt with by exploiting collective wisdom, referring individual cases to a reasonable expectation of privacy. In other words, if one behaves in such a way that one could not reasonably expect to be private, then others are not liable for invading one's privacy. Reasonable expectations change through time and space, making law sensitive to context.

Online, reasonable expectations are diminishing all the time, as our clicks are logged and people generously give information about themselves and others away to their social networks. Surveillance is becoming the norm, with the complicity of many data subjects. But might this be a social harm?

Privacy is essential for the proper functioning of a liberal, democratic society. Some benefits may accrue to the individual (who gains autonomy, a space of intimacy, freedom of speech, and so forth). But equally benefits accrue to society—a free, liberal polity

of autonomous individuals is a public good, in the same way that clean air is. Everyone benefits, even if not everyone contributes.

If privacy is a public, not a private, good, then talking exclusively of rights is not the right way to go. Perhaps we should be talking of the *responsibilities* of privacy too. This would involve something of a culture change, especially in our voyeuristic society.¹ But this would not be unprecedented: it was privacy activists, not the law, which pressured Web sites in the 1990s to respect privacy rather than promiscuously gathering and selling consumers' data.³

Perhaps it is our duty to ensure that reasonable expectations of privacy are kept high. **□**

References

1. Anderson, D. The failure of American privacy law. In B.S. Markesinis, Ed., *Protecting Privacy*, Oxford University Press, Oxford, 1999.
2. Hendler, J. et al. Web Science: An interdisciplinary approach to understanding the Web. *Commun. ACM* 51, 7 (July 2008), 60–69.
3. Hetcher, S.A. *Norms in a Wired World*, Cambridge University Press, Cambridge, 2004.
4. Meiss, M.R., Menczer, F., and Vespignani, A. Structural analysis of behavioural networks from the Internet. *Journal of Physics A: Mathematical and Theoretical* 41, 22 (June 2008); doi: 10.1088/1751-8113/41/22/224022.
5. O'Hara, K. and Shadbolt, N., *The Spy in the Coffee Machine: The End of Privacy As We Know It*. Oneworld, Oxford, 2008.
6. Pitt-Payne, T. Access to electronic information. In C. Reed, and J. Angel, Eds., *Computer Law: The Law and Regulation of Information Technology*, 6th ed., Oxford University Press, Oxford, 2007.
7. Pouillet, Y. and Dinant, J.M. The Internet and private life in Europe: Risks and aspirations. In A.T. Kenyon and M. Richardson, Eds., *New Dimensions in Privacy Law*, Cambridge University Press, Cambridge, 2006.
8. Shadbolt, N., Hall, W., and Berners-Lee, T. The Semantic Web revisited. *IEEE Intelligent Systems* 23, 3 (May/June 2006), 96–101.
9. Walden, I., Privacy and data protection. In C. Reed and J. Angel. *Computer Law: The Law and Regulation of Information Technology*, 6th ed., Oxford University Press, Oxford, 2007.
10. Weitzner, D. et al. Information accountability. *Commun. ACM* 51, 6 (June 2008), 82–87.

Kieron O'Hara (kmo@ecs.soton.ac.uk) is a senior research fellow in Electronics and Computer Science at the University of Southampton, and a research fellow of the Web Science Research Initiative at the University of Southampton, U.K.

Nigel Shadbolt (nrs@ecs.soton.ac.uk) is Professor of Artificial Intelligence and Deputy Head (Research) of the School of Electronics and Computer Science at the University of Southampton, U.K.

Copyright held by author.