Steven M. Bellovin
Columbia University

Policies on Privacy

Privacy is a hotly debated topic. But there isn't just one question—"Should we have more privacy?"—to answer. Rather, there are many, and until we reach consensus on the answers—and their consequences—we cannot agree on what regulation, if any, is appropriate. Bear in mind that the answers can be different for governments and for the private sector, and that this question in particular will be answered very differently by different people or different cultures.

The first set of questions concerns what sort of information can be used. Can an entity obtain information from others about a subject, or is it restricted to information it directly collects? Note that this interacts very directly with the issue of use controls: should collected data be used only for the specified purposes, or can it be repurposed? Secondary use of data—using data for something other than the reason it was originally collected—is one of the biggest sources of privacy problems. This is especially true if multiple datasets are combined.

What, though, constitutes direct collection? If I tag an online picture with someone else's name, is the site entitled to make the association between that person and the picture? Between me and the person I tagged? Between that person and me?

Direct collection is even murkier when it comes to web advertising. Is an on-page advertiser a direct collector? Is it the site hosting the page? Both?

If we want use restrictions, how do we define the categories of uses? What if someone changes his or her mind? Do we want exceptions for, e.g., medical research if identities are protected by contracts?

These questions are common. Two less common issues are the existence of dossiers and the existence, in essence, of time machines.

A dossier is a large compilation of data about a particular individual, similar to what is compiled by credit bureaus and data brokers. These dossiers can be very powerful, but they're what Paul Ohm has referred to as *databases of ruin*. Note, too, that these databases need not contain personally identifiable information to be dangerous; a pseudonymous TiVo account can be just as violative to privacy as one with a real name, since the viewing history can often be deanonymized and linked to a real person.

Dossiers can enable time machines, the ability to see what someone did in the past, before they were of interest to someone else. Governments, of course, love that—but so do marketers. Should such dossiers be allowed to exist? Who should be allowed to query them? Should the information in them "expire" after a while? After how long?

Perhaps, for dossiers, we need revocable anonymity, so that law enforcement can get at the information, but not marketers. That, too, involves a policy decision, albeit a more legalistic one: what are the constraints on police?

It is important for society, not marketers, to answer the questions. For most answers, there are privacy-preserving cryptographic techniques that can at least approximate today's abilities where needed, but without endangering privacy or creating databases of ruin. There are already schemes for things like privacy-preserving targeted ads, verifiable income reporting with anonymous accounts and payment schemes, age verification credentials that don't show a name but are demonstrably valid, and more. I strongly suspect that most other necessary functions can be handled the same way, as soon as the requirements are agreed upon.

There are certainly other important components to privacy, such as a requirement for clear and precise privacy policies by businesses—no more weasel words like *sometimes*, *may*, and *business partners*. But the important thing is to start by making explicit choices about the many different aspects of privacy. ■

Steven M. Bellovin is a professor of computer science and affiliate law faculty at Columbia University. Contact him via https://www.cs.columbia.edu/~smb.

Digital Object Identifier 10.1109/MSEC.2020.2966900
Date of current version: 19 March 2020