S-M-L-XL Data: Big Data as a New Informational Privacy Paradigm

Michael Birnhack*

Can informational privacy law survive Big Data? A few scholars have pointed to the inadequacy of the current legal framework to Big Data, especially the collapse of notice and consent, the principles of data minimization and data specification.[1] These are first steps, but more is needed.[2] One suggestion is to conceptualize Big Data in terms of property:[3] Perhaps data subjects should have a property right in their data, so that when others process it, subjects can share the wealth. However, privacy has a complex relationship with property. Lawrence Lessig's 1999 proposal to propertize personal data, was criticized: instead of more protection, said the critics, there will be more commodification.[4] Does Big Data render property once again a viable option to save our privacy?

To better understand the informational privacy implications of Big Data and evaluate the property option, this comment undertakes two paths. *First*, I locate Big Data as the newest point on a continuum of Small-Medium-Large-Extra Large data situations. This path indicates that Big Data is not just "more of the same", but a new informational paradigm. *Second*, I begin a query about the property/privacy relationship, by juxtaposing informational privacy with property, real and intangible, namely copyright. This path indicates that current property law is unfit to address Big Data.

S-M-L-XL

Context is a key term in current privacy studies. Helen Nissenbaum suggested that in order to evaluate the privacy implications of socio-technological systems, we should ask how these systems affect the informational norms of a given context.[5] This notion fits within the American reasonable expectations test, which indicates whether the

* Professor of Law, Faculty of Law, Tel-Aviv University. Thanks to Omer Tene and Eran Toch for helpful comments and to Dan Largman for research assistance. The research was supported by ISF Grant 1116/12 and Grant 9770-3-873051 of the Israeli Ministry of Science & Technology.

[1] Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239 (2013); Ira S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, INTERNATIONAL DATA PRIVACY LAW 1 (January 25, 2013).

[2] Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63, 64 (2012).

[3] See *e.g.*, Rubinstein, *Big Data*, supra note 1, at 8; OECD, Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value, 220 OECD DIGITAL ECONOMY PAPERS 35 (2013), available at http://dx.doi.org/10.1787/5k486qtxldmq-en.

[4] See LAWRENCE LESSIG CODE AND OTHER LAWS OF CYBERSPACE 159-62 (1999). For criticism, see *e.g.*, Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as an Object*, 52 STAN. L. REV. 1373 (2000). For a complex property/privacy analysis, see Paul M. Schwartz, *Property, Privacy and Personal Data*, 117 HARV. L. REV. 2055 (2004).

[5] HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY AND THE INTEGRITY OF SOCIAL LIFE (2010).

interest in a particular context is worthy of legal protection.[6] Accordingly, I draw a continuum of data contexts, and briefly explore several parameters: the archetypical players, their relationship, the volume, source and kind of data, and the kind of privacy harm that is at stake. For each situation, I note the current legal response.

The continuum is not a neat or rigid classification. The points are indicators of a context. The purpose is to show the development of the contexts, culminating with Big Data. Importantly, the appearance of a new point does not negate or exclude previous points. Big Data raises new challenges, but old and familiar contexts have not elapsed.

Small. The typical Small Data situation assumes one party, usually and individual, that harms another person regarding one informational bit, such as disclosure of a private fact. The data subject and the adversary, to borrow computer scientists' parlance, might have a prior relationship (*e.g.*, family members, neighbors, colleagues), or they are in close proximity: physically (Peeping Tom), socially (a Facebook friend's friend), or commercially (a seller).

Privacy torts developed with small data in mind, and form the common denominator of Warren and Brandies' definition of privacy as the right to be let alone,[7] and Dean Prosser's privacy torts classification.[8] The law attempts to prevent the harm caused to one's need in having a backstage, either physically or mentally. The parties' proximity means that social norms might also be effective.

Medium. Here too there are two parties. The difference is the volume of the data and the balance of power. Unlike the one-time intrusion in the Small Data context, the adversary, now called a data controller, accumulates data and uses it over time. The result is a specified database, created and managed for one purpose, and not further transferred. Typically, the controller is stronger than the subject. Examples are a school that collects data about students, an employer vs. employees, insurance company vs. customers. The technology used can be as simple as a sheet of paper.

In the United States, specific sector-based federal laws apply, *e.g.*, the Family Educational Rights and Privacy Act (FERPA), regulating students' records.[9] The law attempts to assure that the data is not misused. The data controller's legitimate interests are taken into consideration. For example, in the workplace context, the employer has an interest in protecting trade secrets. Thus, the law carves exceptions, limitations, and undertakes various forms of balancing. When the controller is the government, constitutional checks are in operation, under the Fourth Amendment.

[6] Katz v. United States, 389 U.S. 347, 360 (1967).

[7] Samuel Warren & Louis Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890).

[8] William L. Prosser, *Privacy (A Legal Analysis)*, 48 CAL. L. REV. 383, 422 (1960).

[9] 20 U.S.C. §1232g.

Large. As of the 1970s, with technological advancements, it is easier to collect separate bits of data. The volume is much larger, controllers are no longer close to the subjects, and the parties' inequality is enhanced. The paradigmatic situation is a single data collector that processes personal data of many subjects in one database, uses it for multiple purposes, and transfers it to third parties.

Social norms are no longer effective. The concern shifts from the bit to the byte, and then to the megabyte, namely, the database. Once personal data enters a database, the subject can hardly control it. The database contains information without the subject knowing what kinds of data are kept, or how it is used. Moreover, databases may be maliciously abused (nasty hackers, commercial rivals, or enemies), abused to discriminate (by the state, employers, insurance companies, etc.), or reused for new purposes, without the data subject's consent.

The legal response was a new body of law, now called informational privacy or data protection. It assumes that the concentration of the data is dangerous *per se*. Data protection law originated in the 1970s, with the American Ware Report and the Privacy Act of 1974 being an important step,[10] continuing with the influential OECD Guidelines in 1980,[11] and now carried globally by the 1995 EU Data Protection Directive.[12] The common denominator is Fair Information Privacy Principles (FIPPs) that provide data subjects with some (limited) tools to control personal data: notice, consent, limitations on the use of the data, subjects' rights of access and rectification, and the controllers' obligations to confidentiality and data security. In the United States there is a series of sector-based and/or content-based laws that regulate specific contexts. While much of the law is phrased in technologically-neutral language, a close reading reveals that it assumes Large Data.[13]

Extra-Large. Once megabytes turned into terabytes, the risk to personal data shifted yet again. This is Big Data. The volume staggers. There are multiple adversaries. Personal data is gathered from a variety of sources. Data subjects provide a constant stream of accurate, tiny bits of everything they do. It is not always clear who is the data controller. The kind of control also changes. Under Large Data, the way the database was structured mattered. Sensitive kinds of data could be deleted, anonymized, or not collected at all. In contrast, under Big Data, every bit is welcome. The controller does not need to arrange the data at all: all bits are thrown together into one huge bucket. The original context doesn't matter. Bits are constantly collected, taken out of their original context, and mixed. Data is decontextualized only to recontextualize it in a different way. The notion of context-specific laws collapses.

[10] 5 U.S.C. § 552a.

[11] Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980).

[12] Council Directive 95/46, On the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1995 O.J. (L 281) (EC).

[13] Michael D. Birnhack, *Reverse Engineering Informational Privacy Law*, 15 YALE J.L. & TECH. 24, 87-89 (2013).

Current (mostly European) rules would simply prohibit much of XL databases that contain data about identifiable people.[14] Notice and consent per-use are impossible; Big Data operates under a maximization principle rather than a minimization principle. The law breaks down.

Property/Privacy

The property option seems quite tempting. In order to share the wealth, we should be able to protect the wealth in the first place. However, current property law that addresses intangible assets, namely copyright law, does not provide the answer. Here is an outline of the privacy/property juxtaposition along the S-M-L-XL continuum.

S. Property and privacy may overlap. If my home is mine, I can effectively exclude unauthorized intruders and reasonably protect my privacy. The Supreme Court recently concluded that the government's use of drug-sniffing dogs is a "search" under the Fourth Amendment. The Court conducted a property analysis; Justice Kagan's concurrence reached the same conclusion under a privacy analysis.[15] However, privacy and property do not always overlap, as the law protects people, not places.[16]

S., M. From a copyright perspective, for both Small and Medium contexts, the single bit of data does not qualify as proper subject matter. It is an unprotected fact.[17] Neither the data subject nor the controller can rely on copyright law. Without protected property, it is difficult to share the wealth.

L. Real property is irrelevant here. Copyright law may protect the database as a whole, if the selection and arrangement of the facts are original.[18] The individual bits of data remain unprotected. The subject has no property in her personal data, but the data controller might have a property right in the aggregated data. Once the database is protected, there is a reference point for sharing the wealth: it is easier to track down how personal data is processed and used.

XL. Copyright law does not provide the controller with legal protection: the data is not arranged in any particular form, let alone in any original way. Unstructured databases fall outside copyright's subject matter. The controller should seek alternative ways for effective control: the use of technological protection measures is

[14] Article 29 Data Protection Working Party, Opinion 03/2013 on Purpose Limitation (April 2013) at 35, 45-46.

[15] Florida v. Jardines, 569 U.S. __ (2013).

[16] Katz, 389 U.S. at 351.

[17] 17 U.S.C. §102(b).

[18] 17 U.S.C. §103(b); Feist Publications, Inc. v. Rural Telephone Service Company, Inc., 499 U.S. 340 (1991).

one possible avenue, and to the extent that one undertakes reasonable means to keep the data confidential, trade secret law might be another avenue.[19]

*

The continuum of S-M-L-XL data highlights the special characteristics of each data context, the different legal answers, and the ultimate collapse of context under Big Data. Nevertheless, the appearance of Big Data does not mean that previous sizes are eliminated: privacy law is still relevant for the other contexts.

Property law, occasionally suggested as a possible solution for the privacy concerns, is unlikely to offer comfort. Copyright law does not protect the data subject or the controller. Trade secret law might enable the latter effective control, but not assist the data subject.

[19] For an early suggestion, in a Large Data context, see Pamela Samuelson, *Privacy as Intellectual Property?*, 52 STAN. L. REV. 1125 (2000).