Privacy, Ethics, and Analytics

PAUL M.
SCHWARTZ
*University
of California,
Berkeley*

The world is awash in information. In 2003, a UC Berkeley School of Information study found that the amount of new information created every year and stored on media was 5 exabytes, an amount equal to the information stored in 37,000 libraries the size of the US Library of Congress.[1] By 2007, the amount of information stored each year increased to 161 exabytes per year, and growth in the four years since has continued apace.[2] Eric Schmidt, Google's executive chairman, noted that mankind now creates as much information every two days as it had from the dawn of civilization to 2003.[3]

The turn to analytics is a response to this situation. *Analytics* refers to data extraction using efficient, reproducible, and scalable algorithms. The *analytics process* more generally involves data collection, integration, analysis, and decision-making based on this information.

Many organizations use the analytics process to convert data to actionable knowledge. This is the most recent step in *scientific management*, a concept that dates back to the 1880s and the work of Frederick Winslow Taylor.[4] Scientific management has since evolved, replacing managerial instinct with verifiable and documented processes that can be tested and improved. The analytics process now incorporates digital processing to identify patterns in large quantities of information.

The growing popularity of the analytics process has also created new privacy issues. To address these issues, I performed a series of interviews with privacy officers, computer scientists, managers, and lawyers from leading companies, including Acxiom, Google, IBM, and ID Analytics. On the basis of these interviews, I developed a set of high-level ethical principles for private organizations using the analytics process. These ethical principles are intended to represent professional best practices rather than abstract philosophical ideals. Moreover, these standards distinguish among four stages of the analytics process. At various stages of the process, different ethical considerations might be present; therefore, tailoring of rules to meet the challenges of each part of the analytics process is necessary. The goal of these principles is to maximize good results and minimize bad ones for individuals whose information is processed.

The Analytics Process and IT

The analytics process represents a dramatic change in the integration of IT in corporate management. Thomas Davenport and Jeanne Harris, two leading authorities on this technique, describe it as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions."[5] The idea is to convert companies' in-house and third-party digital data to actionable knowledge using scalable processes to accommodate expanding databases.[6]

Until recently, most companies faced significant limitations in their ability to manipulate, process, and learn from data; today, they have access to various software products, ranging from statistical software tools, business intelligence suites, predictive industry applications, and analytic models of enterprise systems, such as those offered by SAP and Oracle. An expanding range of software and enterprise suites is now available to assist the corporate response to the data deluge.

As IT continues to develop, so will the analytics process. A 2009 Forrester report estimated an average compound annual growth of 17 percent for Web analytics and predicted it will be worth US$953 million annually by 2014.[7] The development of IT also means the analytics process's capabilities will evolve and increase. As Davenport and Harris note, "the frontier of decisions that can be treated analytically is always moving forward."[5] Examples of these emerging capabilities include analytics in the cloud, real-time analytics, and navigation and analysis of unstructured information through natural language processing.

The analytics process allows data to drive corporate decision-making, documenting opaque processes and replacing conventional wisdom—when wrong—with tested approaches. As one CEO put it, "In God we trust; all others bring data."[8]

Four Stages of the Analytics Process

The analytics process has four stages—*collection*, *integration and analysis*, *decision-making*, and *review and revision*. Each one can raise different privacy issues.

- Collection is the stage at which information is assembled. Typically, companies approach the collection process in a broad fashion. The identification of valuable, interesting, or unexpected data patterns might depend on access to a wide sweep of data. Indeed, companies frequently decide whether information will be relevant to their decision-making on the basis of analytics. Thus, the collection stage is a necessary part of this framework because identifying connections and the meaning of different variables before starting the analytics process is often impossible.
- In the integration and analysis stage, companies assess the information at hand and execute analytics. Integration requires aggregating data from multiple sources inside and outside an organization[9] as well as standardizing subjects for inclusion and taking steps to ensure data quality. For example, the Partners for Child Passenger Safety established a special database based on a subset of claims about accidents involving children in different jurisdictions.[10] In this study, researchers trained claim representatives in 365 field offices, who used an electronic insurance claims database to

gather data on children involved in automobile accidents. Their goal was to determine the risks of "premature graduations of young … children to seat belts from child restraint systems."[10] The analysis stage then entails examining the results for patterns, which might lead companies to gather more information and perform additional integration and analysis.

- The decision-making stage occurs when companies act on the results of the integration and analysis stage. For example, Netflix makes real-time suggestions for films based on its Cinematch analytics, and financial services entities use analytics to understand their customers and prospects, and to target appropriate products and services to them at the right moment. Doing so lets companies market to individuals in a highly dynamic fashion. It also means that companies can access—sometimes in real time—highly sensitive personal information about individuals' likes and dislikes.
- Successful companies continue to review and revise their analytics model. Businesses should use an analytics process that works not only today, but also in the future. Business intelligence software codifies a set of assumptions to forecast and optimize their products and services. However, as Kenneth Bamberger warns, these choices "may be embedded in a way that is difficult to identify or alter as contexts change."[11] Companies should verify results to prove that the analytics process does, in fact, lead to better decisions. For instance, Amazon shares information about its analytics process with each customer to let him or her know why it makes certain recommendations. Customers can manage their

browsing history by turning off Amazon's ability to track it or by deleting items in it. Amazon also refines recommendations by using rating systems and letting customers exclude certain items, such as gifts, from future recommendations. This transparency allows customer input, gives Amazon valuable feedback, and heightens consumer trust. Such verification is important to avoid the dangers that can arise from mistaken assumptions, bad programming, or other factors.

These four stages don't always occur in sequence. For example, the first two stages might repeat several times before a company decides to make decisions on the basis of the analytics results. Several cycles of decision-making might also occur before a company decides to review and revise its model for the analytics process.

In addition, different international jurisdictions are likely to reach different conclusions about whether a collection of information implicates legal regulations concerning processing. In particular, the EU Data Protection Directive has an expansive definition of the concept of the "processing of personal data."[12] In the US, data collection might not necessarily be considered as triggering analogous legal safeguards. As a result, this paper's use of terms such as *collection* and *processing* does not coincide with definitions in the EU Data Protection Directive and other EU documents.

The Ethical Use of the Analytics Process

Here, I present general rules for the ethical use of the analytics process during these four stages.

Overarching Ethical Standards

The following standards apply to all four stages of the analytics process:

- Companies should comply with legal requirements when using the analytics process.
- Companies should assess, beyond

The law considers children especially vulnerable to outside influences. Social science research has shown that young children have

Digital marketing platforms directed toward children have increased dramatically, and the privacy and security of their data deserves special consideration. Companies should take this into account.

legal requirements, whether their use of the analytics process reflects cultural and social norms about acceptable activities.
- Companies should assess the impact of their use of the analytics process on the basis of various stakeholders' trust in the company. Relevant stakeholders include consumers, other businesses, government, and nongovernmental policymakers.
- Companies should use accountable measures in the analytics process. Accountability begins with an acknowledgment that this technique can have both negative and positive impacts on individuals. Companies should develop organizational policies that govern information management and personnel training, and should designate individuals to oversee data processing operations and decision-making. Accountable measures should be appropriately tailored to specific, identified risks of the analytics process.
- Companies should implement appropriate safeguards to protect the security of information they use in this process. Data security should be reasonable when measured against the kind of information that's collected and processed—and the decisions made with it.
- Companies should assess whether their use of the analytics process involves sensitive areas and, if so, accompany it with safeguards proportionate to the risk.

difficulty discriminating between advertising and programming.[13] It also suggests that teenagers are especially vulnerable to targeted marketing, owing to the pressures of adolescence.[14] Therefore, a company's use of analytics to shape contacts with children raises special ethical concerns.

At the same time, analytics involving information about children can be highly beneficial, leading to medical advances and public safety breakthroughs. So, a simple ban on using the analytics process on information about children would be counterproductive. Nonetheless, digital marketing platforms directed toward children have increased dramatically, and the privacy and security of their data deserves special consideration.[15] Companies should take this into account and place responsible limits on their use of the analytics process in this area.

Stage-Specific Ethical Standards

The following are ethical considerations related to each of the four stages.

Collection. Companies should avoid collecting certain kinds of information for use in the analytics process. In making this judgment, companies should consider legal, cultural, and social factors as well as risks to themselves and affected individuals. For ex-

ample, a wide group of industry associations in the US has agreed that entities shouldn't collect and use an individual's pharmaceutical prescriptions or medical records for online behavioral advertising without consent.[16] The Draft EASA (European Advertising Standards Alliance) Best Practice Recommendation goes considerably further,[17] incorporating the EU Data Protection Directive's definition of personal data, which extends to "revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and ... data concerning health or sex life."[18] Under the EASA recommendations, companies need a Web user's explicit consent before creating online behavioral advertising based on these categories.

Integration and analysis. As companies assess information and execute analytics, they face a different set of ethical obligations. Considerations include whether data is of sufficient quality for its likely future use, and whether responsible choices will be made regarding the need to gather additional information and engage in additional integration and analysis. Companies should also anonymize personal information, when appropriate, during analysis.

Decision-making. When acting on analytics results, companies should use analytical output that is reasonably accurate based on the nature and significance of the underlying decisions. For more important decisions with greater impact on individuals, they should rely on more accurate data. Companies should also offer reasonable compensatory controls, when appropriate, and develop reasonable mitigation processes and remedies when the analytics process leads to decisions that harm individuals.

Finally, companies should assess whether their decision-making reflects legal, cultural, and social norms about acceptable activities and take steps to comply with these norms.

Review and revision. Companies should consider the following to develop a sustainable system:

- Engage in ongoing review and revisions of their use of the analytics process.
- Ensure that personal information is reasonably relevant and accurate for their purposes.
- Be responsive to the impact and unforeseen consequences of their decisions that raise ethical questions.
- Use only predictive information in the analytics process and revise procedures to exclude non-predictive information when reasonable and appropriate.

An operational framework for these standards will depend on each company's discrete managerial, regulatory, and ethical issues.

When using the analytics process, companies should consider the risks it poses to individuals' information privacy as well as develop responsible measures to accompany its use. This set of ethical standards calls on companies to adopt accountable approaches that reflect the specific risks in a given use of the analytics process. □

References

1. P. Lyman and H.R. Varian, *How Much Information?* Univ. of California, Berkeley, 2003; www2.sims.berkeley.edu/research/projects/how-much-info-2003.
2. S. Gaudin, "The Digital Universe Created 161 Exabytes of Data Last Year," *InformationWeek*, 7 Mar. 2007; www.informationweek.com/news/internet/search/showArticle.jhtml?articleID=197800880.
3. G. Seigler and E. Schmidt, "Every 2 Days We Create as Much Information as We Did up to 2003," *TechCrunch*, 4 Aug. 2010; http://techcrunch.com/2010/08/04/schmidt-data.
4. F.W. Taylor, *Principles of Scientific Management*, Dover, 1911.
5. T.H. Davenport and J.G. Harris, *Competing on Analytics*, Harvard Business School Press, 2007.
6. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed., Morgan Kaufmann, 2005.
7. J. Lovett, "US Web Analytics Forecast, 2008 to 2014: The Future Brings a Shift in Role," Forrester Research, 27 May 2009; www.forrester.com/rb/Research/us_web_analytics_forecast%2C_2008_to_2014/q/id/53629/t/2.
8. T.H. Davenport et al., "Data to Knowledge to Results," *California Management Rev.*, vol. 43, no. 2, 2011, pp. 117–138.
9. T.H. Davenport, J.G. Harris, and R. Morison, *Analytics at Work*, Harvard Business Press, 2010.
10. F.K. Winston et al., "The Danger of Premature Graduation to Seat Belts for Young Children," *Pediatrics*, vol. 105, no. 6, 2000, pp. 1179–1183.
11. K.A. Bamberger, "Technologies of Compliance: Risk and Regulation in a Digital Age," *Texas Law Rev.*, vol. 88, no. 4, 2010, p. 669.
12. "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data," art. 2(b), 1995; http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf.
13. V.C. Strasburger, B.J. Wilson, and A.B. Jordan, *Children, Adolescents, and the Media*, 2nd ed., Sage, 2009.
14. J.L. Harris, K.D. Brownell, and J.A. Bargh, "The Food Marketing Defense Model: Integrating Psychological Research to Protect Youth and Inform Public Policy," *Soc. Issues & Policy Rev.*, vol. 3, no. 1, 2009, pp. 211–271.
15. J. Chester and K. Montgomery, "Interactive Food and Beverage Marketing: Targeting Children and Youth in the Digital Age," Berkeley Media Studies Group, May 2007; http://digitalads.org/documents/digiMarketingFull.pdf.
16. "Self-Regulatory Principles for Online Behavioral Advertising" 17 Jul. 2009; www.iab.net/public_policy/behavioral-advertising principles.
17. O. Gray and A.M. Wade, "Draft EASA Best Practice Recommendation," 7 Dec. 2010; http://ec.europa.eu/justice/news/consulting_public/0006/contributions/organisations/epc_annex2b_en.pdf.
18. *EU Directive 95/46/EC—The Data Protection Directive*, art. 8(1); www.dataprotection.ie/viewprint.asp?fn=/documents/legal/6aii-2.htm.

Paul M. Schwartz *is a professor of law at University of California, Berkeley, and a director of the Berkeley Center for Law and Technology. Contact him at pschwartz@law.berkeley.edu.*