



PRIVACY & SECURITY LAW



REPORT

Reproduced with permission from Privacy & Security Law Report, 10 PVLR 70, 01/10/2011. Copyright © 2011 by The Bureau of National Affairs, Inc. (800-372-1033) <http://www.bna.com>

Data Protection Law and the Ethical Use of Analytics



By PAUL M. SCHWARTZ

I. Analytics: An Introduction

Organizations now work in a data-rich environment. As the Article 29 Working Group of the EU recently noted, “[W]e are witnessing a so-called ‘data deluge’ effect, where the amount of personal data that exists, is processed and is further transferred continues to grow.”¹ From all indications, the data deluge will not only continue, but increase.

¹ Article 29 Data Protection Working Party, Opinion 3/2010 on the Principle of Accountability 4 (July 13, 2010).

Professor of Law, Berkeley Law School, University of California, Berkeley; Director, Berkeley Center for Law & Technology.

In 2003, a study at the UC Berkeley School of Information found that the amount of new information being created every year and stored on media was 5 exabytes.² That amount is equal to the information stored in 37,000 libraries the size of the Library of Congress in the United States. By 2007, however, the amount of information stored each year had increased to 161 exabytes a year.³ This development has continued apace. In 2010, Google CEO Eric Schmidt noted that mankind now creates as much information every two days as it had from the dawn of civilization to 2003.⁴

The turn to analytics is a response to this situation. Analytics involve the use of statistics, algorithms, and other tools of mathematics, harnessed through information technology, to use data to improve decision-making. A wide variety of organizations use analytics in their operations. Analytics are used by government, for example, but this white paper concentrates on how private-sector organizations use this technique.⁵ It does

² Peter Lyman & Hal R. Varian, *How Much Information? 2003* (University of California, Berkeley 2003), <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>.

³ Sharon Gaudin, *The Digital Universe Created 161 Exabytes of Data Last Year*, INFORMATIONWEEK (Mar. 7, 2007), <http://www.informationweek.com/news/internet/search/showArticle.jhtml?articleID=197800880>.

⁴ MG Seigler, *Eric Schmidt: Every 2 Days We Create as Much Information as We Did up to 2003*, TECHCRUNCH (Aug. 4, 2010), <http://techcrunch.com/2010/08/04/schmidt-data/>.

⁵ For a White Paper evaluating the use of analytics by government in Canada to detect fraud, see Adam Kardash & Ruth Belcher, *Privacy No Obstacle in Implementation of Risk-based Fraud Detection Solutions in the Public Sector When Deployed with Suitable Controls*, ACCESS PRIVACY (2010), available at

so because distinctive regulatory and ethical issues are likely to arise for different categories of enterprises.

This white paper offers a contextual examination of analytics and develops ethical standards for private organizations using this technique. The term “contextual” is used here in reference to an organization’s need to consider the risks that a specific application of analytics poses to privacy and the kind of responsible processes that should accompany the use of analytics generally. This white paper finds that analytics tend to be applied to four stages of a data life-cycle: (1) collection, (2) integration and analysis, (3) decision-making, and (4) review and revision.

Its ethical standards for private organizations using analytics were developed through a series of interviews and discussions with the leading companies that participated in this project of the Centre of Information Policy Leadership at Hunton & Williams. The resulting standards acknowledge that analytics can have a negative as well as a beneficial impact on individuals. Thus, the white paper requires implementation of accountable processes that are tailored to the specific, identified risks of analytics used. The guidelines further require development of organizational policies that govern information management and training of personnel. A company should also place responsibility for data processing operations and decision on designated individuals within the company. The following report is an abridged version of the full white paper, which is available online.⁶

A. Why Analytics?

Analytics provide a way for organizations to draw on the great quantities of information in their control or available from third parties and to use the data to make better decisions and to create new products and services. In the definition of Thomas Davenport and Jeanne Harris, two leading authorities on this technology, analytics refers to “the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions.”⁷ The idea is to take the information that entities have, or to which they can gain access, and to convert it to actionable knowledge.⁸ This approach is now popular in the corporate world. As a blogger on the Harvard Business Review website concisely observed in September 2010, “Analytics are now king.”⁹

Analytics represent a dramatic change from the old approach to corporate management. In the past, many corporate decisions, even the majority in some settings, were undocumented and managed by executives

through reliance on their instincts. Despite all the information available, most companies faced significant limitations with their ability to manipulate, process, and learn from data. Today, the analytical process permits the decision-making process to be driven by the use of data. It seeks to document opaque processes and to replace conventional wisdom, if wrong, with tested approaches. As one CEO put it, “In God we trust; all others bring data.”¹⁰

B. The Role of Information Technology

Information technology (IT) has been a powerful motor in driving changes in the use of analytics. Companies now have access to in-house and third-party digital data and different kinds of software products, ranging from statistical software products, business intelligence suites, predictive industry applications, and analytical modules of major enterprise systems, such as offered by SAP and Oracle.¹¹ An expanding range of software and enterprise suites are now available to assist the corporate response to the data deluge.

IT continues to develop, which means that the use of analytics will grow. Regarding analytics as a growth business, a Forrester Report from 2009 estimates an average compound annual growth of 17 percent alone for web analytics.¹² It predicts that this branch of analytics will be a \$953 million annual business by 2014. The development of IT also means that the capabilities of analytics will evolve and increase. As Davenport and Harris express this idea, “The key message is that the frontier of decisions that can be treated analytically is always moving forward.”¹³ Examples of these emerging capabilities include analytics in “the cloud,” real-time analytics, and the navigation and analysis of unstructured information through “natural language processing, search, inference and categorization.”¹⁴

C. The Different Stages of Analytics

In studying analytics, one can identify four discrete stages in their use. These steps are: (1) collection; (2) integration and analysis; (3) decision-making, and (4) review and revision. Each period of analytics raises different kinds of issues for privacy.

Two further points are necessary. As an initial matter, these steps do not always occur in sequence from the first to the fourth stage. For example, the first two stages may repeat several times before a company decides to make decisions based on the results from analytics. As a further initial caveat, the use of terms such as “collection” and “processing” below may not necessarily fit within the definitions of these terms in the EU Data Protection Directive and other EU documents. In particular, there is a potential for diverse legal conclusions in different international jurisdictions about whether a “collection” of information implicates legal regulations concerning “processing.” For example, the EU Data Protection Directive has an expansive definition of the concept of the “processing of personal

http://www.accessprivacy.com/docs/paper_privacy-no-obstacle.pdf.

⁶ Paul M. Schwartz, *Data Protection Law and the Ethical Use of Analytics* (2010), at http://www.huntonfiles.com/files/webupload/CIPL_Ethical_Underpinnings_of_Analytics_Paper.pdf.

⁷ THOMAS H. DAVENPORT & JEANNE G. HARRIS, *COMPETING ON ANALYTICS* 7 (2007).

⁸ As Thomas Davenport and co-authors explain, “The analytic process makes knowledge from data.” Thomas H. Davenport et al., *Data to Knowledge to Results*, 43 CAL. MGMT. REV. 117, 128 (2001).

⁹ Michael Fertick, *Hire Great Guessers*, HARVARD BUSINESS REVIEW BLOG (Sept. 2, 2010 8:30 AM), http://blogs.hbr.org/cs/2010/09/hire_great_guessers.html.

¹⁰ Davenport et al., *Data to Knowledge to Results*, at 136.

¹¹ DAVENPORT & HARRIS, *COMPETING ON ANALYTICS*, at 7–8.

¹² John Lovett, U.S. WEB ANALYTICS FORECAST, 2008 TO 2014: THE FUTURE BRINGS A SHIFT IN ROLE (Forrester May 27, 2009).

¹³ DAVENPORT & HARRIS, *COMPETING ON ANALYTICS*, at 14.

¹⁴ *Id.*

data.”¹⁵ In the United States, the collection of data at an initial stage of analytics may not necessarily be considered as triggering analogous safeguards.

At this juncture, this white paper will set out the different steps that generally occur in the process of analytics. First, collection refers to the stage at which information is assembled. Typically, companies seek to approach the collection process in a broad fashion. This approach occurs because it is frequently not possible to identify connections and the meaning of different variables before one starts the analytic process. As Davenport, Harris, and Morison describe the ideal approach to collection, it is for a company “to tap into and exploit data that no one else has.”¹⁶

Second, integration and analysis is the stage at which companies assess the information at hand and execute analytics. Data integration requires “the aggregation of data from multiple sources inside and outside an organization.”¹⁷ As an example, the Partners for Child Passenger Safety has established a database of claims from an insurance company about accidents involving children in different jurisdictions.¹⁸ Analysis is the process of examining the results for patterns and results. This assessment may also coalesce around a choice to gather additional information and additional integration and analysis.

Third, the decision-making stage occurs when companies act on the results of the analytics. For example, Amazon will make real-time recommendations to a consumer for one book and not another based in its analytics. Netflix similarly makes real-time suggestions for films based on its Cinematch analytics. Financial service entities use analytics to understand their customers and prospects and to target appropriate products and services at the right moment. Doing so permits marketing to individuals in a highly dynamic fashion.

Fourth, a firm will review and revise its analytics. Businesses should seek to have 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, but, as Kenneth Bamberger has warned, these choices “may be embedded in a way that is difficult to identify or alter as contexts change.”¹⁹ Towards a similar end, Davenport and Harris note that the companies who make the best use of analytics, whom they term “analytical competitors,” are ones that are focused on “continuous analytics re-

newal.”²⁰ They recognize the need for “carefully monitoring outcomes and external conditions to see whether assumptions need to be modified.”²¹ Amazon is one of the companies that has opened up its analytics process, and let customers know why it has made a certain kind of recommendation. This transparency permits customer input and provides Amazon with valuable feedback. It also heightens consumer trust in Amazon. In sum, firms should verify results in order to prove over time that the analytics do, in fact, lead to better decisions. This process is important for business reasons and also because of the dangers that can flow from mistaken assumptions, bad programming, or other factors.

II. The Ethical Use of Analytics

In this Part, the White Paper develops rules for the issue of the ethical use of analytics during these four discrete stages. Some elements of the ethical use of analytics apply at all stages, however, and will be discussed in this Part as an initial matter. In each of the following sections, I summarize the key issues through bullet points.

A. Overarching Ethical Requirements

The ethical use of analytics should be driven by a company’s assessment of the impact of legal, cultural, and other factors on its obligation to be a socially responsible entity.

- A company should comply with legal requirements in its use of analytics.
- A company should assess, beyond legal requirements, whether its use of analytics reflects cultural and social norms about acceptable activities.
- A company should assess the impact of its use of analytics on the trust in the company held by a wide range of stakeholders. Relevant stakeholders can include consumers, other businesses, government, and non-governmental policymakers.
- A company should use analytics through accountable processes. Accountability begins with an acknowledgment that analytics can have a negative as well as beneficial impact on individuals. A company should also develop internal policies that center on forward-looking rules of information management and training of personnel. Accountable processes for analytics should be appropriately tailored to counter the risks raised by specific uses of analytics.
- A company should implement appropriate safeguards to protect the security of information that it uses in analytics. Data security should be reasonable when measured against the kind of information that is collected and processed, and the decisions that are made with it.
- A company should assess whether its use of analytics involves sensitive areas and, if so, accompany it with reasonable safeguards proportionate to the risk.

Firms should also make ethical decisions when they use analytics in areas that affect children. To be sure, it can be highly beneficial to carry out analytics involving information about children. In particular, analytics involving aggregate data sets can lead to significant medical advances and public safety breakthroughs. As a re-

¹⁵ See 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 O.J. (L 281) (“‘processing of personal data’ shall mean any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaption, or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction”) [hereinafter Data Protection Directive].

¹⁶ THOMAS H. DAVENPORT, JEANNE G. HARRIS & ROBERT MORISON, *ANALYTICS AT WORK* 25 (2010).

¹⁷ *Id.* at 29.

¹⁸ Flaura K. Winston et al., *The Danger of Premature Graduation to Seat Belts for Young Children*, 105 *PEDIATRICS* 1179, 1179 (2000).

¹⁹ Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 *TEX. L. REV.* 669, 676 (2010).

²⁰ DAVENPORT & HARRIS, *COMPETING ON ANALYTICS*, at 109.

²¹ *Id.* at 129.

sult, a simple ban on analytics about children would be highly counter-productive. Nonetheless, there has also been a dramatic increase in digital marketing platforms directed towards children, and the privacy and security of their data deserves special consideration.²² Due to the special vulnerability of children, the use of analytics to shape contacts with them raises special ethical concerns.

- A company should take into account the special vulnerability of children in placing responsible limits in its use of analytics.

B. Stage One: Collection

- A company should not collect certain kinds of information for use in analytics. Its analysis should be based on legal, cultural and social factors. In making this judgment, an ethical company should also consider risks to the company and affected individuals.

C. Stage Two: Integration and Analysis

After collection, a company will assess the information at hand and execute the analytics. At this stage, the company faces a different set of ethical obligations.

- Companies should anonymize personal information when appropriate in their analysis of it.

D. Stage Three: Decision-making

The decision-making stage occurs when companies act on the results of the analytics.

- A company should engage in decision-making based on analytical output that is reasonably accurate, based on the nature and significance of the underlying decisions. If it seeks to reach decisions that are more important and of a higher impact for the individual, it should rely on data of a greater accuracy.
- A company should make available reasonable compensatory controls when appropriate.

- A company should develop reasonable mitigation processes and reasonable remedies as appropriate when analytics lead to decisions that harm individuals.
- A company should assess whether its decision-making with analytics reflects legal, cultural, and social norms about acceptable activities and take steps, when needed, to comply with these norms.

E. Stage Four: Review and Revision

Finally, a company should review and revise its analytics as part of developing a process that works not only today, but in the future.

- Companies should engage in ongoing review and revisions of their use of analytics.
- Companies should review and revise analytics to make sure that personal information will be reasonably relevant and accurate for the purposes for which they are used.
- Companies should be responsive to the impact of decisions and unforeseen consequences of analytics that raise ethical questions.
- Based on their review and revision, companies should only use information that is predictive in analytics and revise procedures, when reasonable and appropriate, to exclude non-predictive information.

VI. Conclusion

This white paper has explored different dimensions of analytics for information privacy. This technique for data use, which is growing in its rate of adoption and expanding its overall capacities, brings with it the potential for positive and negative effects. In response, the white paper has argued for a contextual examination of analytics. Organizations should consider the risks that a specific use of analytics poses to privacy and develop responsible processes to accompany its use. This project has also identified four different stages of analytics and argued that responsible processes should be tailored to each step. It has developed a set of ethical standards for the use of this technique and called upon companies to adopt accountable processes that reflect the specific risks in a given use of analytics.

²² See JEFF CHESTER & KATHRYN MONTGOMERY, BERKELEY MEDIA STUDIES GROUP, *INTERACTIVE FOOD & BEVERAGE MARKETING: TARGETING CHILDREN AND YOUTH IN THE DIGITAL AGE* (May 2007).