# DATA PROTECTION LAW
# AND
# THE ETHICAL USE OF ANALYTICS

By Paul M. Schwartz

THE CENTRE
FOR INFORMATION
POLICY LEADERSHIP
HUNTON & WILLIAMS LLP

## ACKNOWLEDGEMENTS

The Centre for Information Policy Leadership would like to express its appreciation to IBM for organizing and hosting a workshop where the issues addressed in this paper were discussed. It is grateful to all of the participants in this project for taking part in the workshop and for their insights and assistance. The participants include:

Acxiom Corporation
Bank of America
Google
Heenan Blaikie LLP
Hewlett-Packard Company
IBM Corporation
ID Analytics
IMS Health Incorporated
Intel Corporation
Intuit Inc.
LexisNexis Group
MasterCard
Microsoft Corporation
TURN
Yahoo! Inc.

# DATA PROTECTION LAW AND THE ETHICAL USE OF ANALYTICS

**By Paul M. Schwartz***



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

# EXECUTIVE SUMMARY

The term “analytics” refers to the use of information technology to harness statistics, algorithms, and other tools of mathematics to improve decision-making. A wide variety of organizations use analytics to convert data to actionable knowledge. Analytics represent a change from the longstanding approaches to management that often rely on instinct and largely are unsupported and undocumented. Analytics permit corporate decision-making to be driven, assessed and tested by the use of data.

This paper offers a contextual examination of analytics. 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.

The white paper further discusses how rules about the collection and processing of personal information can reflect different legal, social, and cultural values in different countries. These disparities raise considerable challenges to international companies that work in a variety of jurisdictions. The differences in background values can also raise challenges to a distributed business strategy that involves partnering with other entities on a global basis.

Finally, the paper proposes ethical standards for private organizations using this technique. These guidelines were developed through a series of interviews and discussions in a workshop with the leading companies that participated in this project. The 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 polices 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 ethical standards, as applied generally and to different stages of data life-cycle identified in this paper, are as follows:

## Overarching requirements

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

### Stage One: Collection

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

### 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 refrain from use of information once integration and analysis show it to be of insufficient quality for the intended purpose.
- Companies should anonymize personal information when appropriate in their analysis of it.

### 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 decision. If it seeks to reach decisions that are more important and of a higher impact for the individual, it should rely on data of 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.

### 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 the 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.

# 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.

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

In this part, the white paper discusses the rise of analytics. This technique allows organizations to draw on the great amounts of information that are now available and allows decision-making to be data driven. It is also greatly shaped by developments in IT, and is a growing field in terms of both its rate of adoption and the new kinds of capabilities that it is gaining.

This white paper's Part II examines the need for a contextual examination of analytics. The term "contextual" refers to an organization's need to consider the risks that a specific use of analytics poses to privacy and the kind of responsible processes that should accompany their use. Part II also traces important changes in the environment of data-protection law. These include a professionalization of information management at leading companies. Moreover, data-protection law itself has never been static. In the 21st century, in particular, there have been important discussions about the "modernization" of this law and the introduction of principles of accountability and proportionality. This discussion offers important insights for the ethical use of analytics.

In Part III, the white paper considers different examples of analytics in action. These include multichannel marketing; fraud prevention and data security; health care research; and a variety of products for direct use by individuals, such as financial software, influenza tracking software, and translation software. The white paper also identifies four distinct stages of analytics and argues that responsible data processing should be tailored to the discrete stages in which analytics are used.

Part IV looks at the complex questions that analytics raise for data-protection law built around traditional Fair Information Practices (FIPs) as expressed in important international guidelines. Particular areas of tension are

[1] Article 29 Data Protection Working Party, Opinion 3/2010 on the Principle of Accountability 4 (July 13, 2010).
[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=197800.
[4] MG Seigler, *Eric Schimdt: 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 http*://www.accessprivacy.com/docs/paper_privacy-no-obstacle.pdf.

the existing rules about (1) "automated individual decisions" and (2) the set of related rules about "purpose specification" and "use limitation." These international rules reflect, to some extent, legal, social, and cultural values that may be somewhat different than regulations for similar areas in the United States. The challenge for private-sector organizations operating in different countries, and in an age of international data flows, will be to develop ethical standards that will permit a harmonization of their global operations.

Finally, Part V of the white paper proposes a set of ethical guidelines for the use of analytics. The white paper's ethical standards have been developed through a series of interviews and a workshop with experts from a cross-section of private sector companies. The organizations participating in this project include a variety of leading companies that currently use analytics.

## 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."[6] The idea is to take the information that entities have, or to which they can gain access, and to convert it to actionable knowledge.[7] 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."[8]

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."[9]

## 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.[10] An expanding range of software and enterprise suites are now available to assist a 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.[11] 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

[6] Thomas H. Davenport & Jeanne G. Harris, Competing on Analytics 7 (2007).
[7] 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).
[8] 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.
[9] Davenport et al., *Data to Knowledge to Results*, at 136.
[10] Davenport & Harris, Competing on Analytics, at 7–8.
[11] John Lovett, U.S. Web Analytics Forecast, 2008 to 2014: The Future Brings a Shift in Role (Forrester May 27, 2009).

moving forward."[12] Examples of these emerging capabilities include analytics in "the cloud," real-time analytics, and the navigation and analysis of unstructured information.

Cloud computing is the location of computing resources on the Internet in a way that makes them highly dynamic and scalable. Data transmissions now occur as part of a networked series of processes and on a global scale. This distributed computing environment permits dramatic use of analytics in processing decisions. Today, computing activities can be shifted among servers in different countries depending on load capacity, personnel availability, time of day, and other factors.

Greater computing ability also has meant more real time analytics.[13] Netflix uses Cinematch, its proprietary movie-recommendation software, to: (1) define so-called "movie clusters," which are films with similarities; (2) link customer movie rankings to these films; (3) analyze thousands of ratings per second, and (4) factor in a customer's current website behavior. The end result? Cinematch generates a real-time personalized webpage for each Netflix customer.[14] Recommendations of movies on the resulting individual webpage are made both to fit the customer's taste and to optimize Netflix's inventory.[15] Cinematch has been found to steer sixty percent of Netflix rentals.[16] It is able to help customers find older movies and independent films and, as the *New York Times* stated, encourage them "to consume more stuff."[17]

As for the use of unstructured information, computer science continues to improve its ability to search through different kinds of data sets. Bruce McCabe has observed, "Most new digital information exists in the form of text, images, audio and video that has little structure or organisation."[18] More of this information is now available for computer analysis due to "natural language processing, search, inference and categorization."[19]

[12] Davenport & Harris, Competing on Analytics, at 14.
[13] *Id*. at 176–177.
[14] *Id*. at 4.
[15] *Id*.
[16] Clive Thompson, *If You Liked This, You're Sure to Love That*, N.Y. Times, Nov. 23, 2008, *available at* http://www.nytimes.com/2008/11/23/magazine/23Netflix-t.html?pagewanted=print.
[17] *Id*.
[18] Bruce McCabe, *The Future of Business Analytics,* S2 Intelligence, May 2007, at 3.
[19] *Id*.

## II. ANALYTICS, DATA PROTECTION LAW, AND ETHICS

In this part, the white paper discusses reasons why it is difficult to establish a line—in advance—between analytics that raise data protection concerns and those that do not. Since this judgment cannot be made in an abstract fashion, analytics should be evaluated in a contextual fashion. This part will also consider recent changes in the regulatory dialogue about data protection law. Discussions about the need for "modernization" of this area of law have concerned the role of companies and the twin concepts of "accountability" and "proportionality." Drawing on this dialogue, later sections of this white paper will propose ethical standards for companies to use analytics in accountable and proportional ways.

### A. Data Everywhere: The Need for Contextual Examination of Analytics

As an initial point, it is difficult *a priori* to identify uses of analytics that fall inside or outside data protection concerns. The first reason for this failure of abstract analysis is that the dividing line between personally identified information (PII) and non-PII can be difficult to trace. The second reason is that the use of analytics is a highly dynamic process and can easily create privacy implications where none were originally predicted. As a consequence, the implications of analytics for data protection should be assessed through a contextual analysis that evaluates risks involved in a specific application. Note as well, however, that some data sets will remain aggregate and non-specific as to any person, and will not implicate data protection concerns.

Regarding the question of PII versus non-PII, recent work in computer science has shown how easy it can be to trace non-PII to identifiable individuals. Latanya Sweeney has demonstrated that combining a zip code, birth date, and gender is sufficient to identify eighty-seven percent of the U.S. population.[20] As a further example, a study involving Netflix movie rentals was able to identify eighty percent of people in a supposedly anonymous database of 500,000 Netflix users; the identification was triggered by their ratings in the Netflix database of at least three films.[21] In this study, Arvind Narayanan and Vitaly Shmatikov were able to link the information in the Netflix data sample to movie ratings that the same individuals gave to films in the Internet Movie Database (IMDb). Paul Ohm has drawn on this research and other studies to argue, "No matter what the data administrator does to anonymize the data, an adversary with the right outside information can use the data's residual utility to reveal other information."[22]

In addition to the increasingly indistinct line between PII and non-PII, initial uses of analytics can frequently be modified, sometimes only slightly, to create privacy issues. In other words, analytics can begin in a "safe" category—that is, one that does not involve informational privacy concerns—but this process can easily be modified or otherwise developed so that it does, in fact, implicate privacy.

To illustrate, we can consider the use of analytics by a package delivery service to improve its routes, a hospital in stocking medical supplies, a retail outlet in selecting the types of blinds to sell, and a bank in choosing which branches to keep open and which to close.[23] These might all seem, at the start, to fall into that "safe" category. Yet in these scenarios, the companies involved may develop analytics, over time, to drill down in data sets and identify specific individuals. The delivery service will want to know why certain routes are less profitable and why others are more profitable. A single customer on the route may be driving its profitability. In that case, the business will want to know how to personalize and improve the service for this individual. The hospital's analysis

[20] Latanya Sweeney, *Uniqueness of Simple Demographics in the U.S. Population* (Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-WP4, 2000).
[21] Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets* (University of Texas at Austin 2008), http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf.
[22] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. Rev. 1701, 1752 (2001).
[23] *See generally* Thomas H. Davenport, Jeanne G. Harris & Robert Morison, Analytics at Work (2010).

of inventory levels will likely lead to an examination of specific patterns of utilization of supplies by physicians, nurses, and staff. The analysis by the retail outlet of its sales of blinds will generate information about which salesmen are most effective with different customer cohorts.

As for the bank, in its decisions about which branches to keep open and which to close, it will go beyond aggregate customer transactional data. The key question for this business, as for all others, ultimately regards the satisfaction of its individual customers. It may be possible, for example, for this financial institutional to customize its products and services for customers. In short, these examples concerning the organization of supply chains, the improvement of delivery routes, the optimization of inventories, and the location of branch offices can turn to questions that reach down to the individual.

As a final example, we can consider sports analytics. Today, sports teams in the United States, Europe, and elsewhere make significant use of analytics in building their roster, in contract negotiations with players, and in on-the-field strategic decisions. MONEYBALL (2003), by Michael Lewis, is a best-selling account of this technique. Lewis shows how Billy Beane, the general manager of the Oakland Athletics baseball team, used rigorous statistic analysis to field the best team at the lowest cost. Through the use of analytics, Beane was able to draw on modern empirical tools for judging player performance rather than the conventional, and sometimes incorrect, wisdom of baseball lore.[24]

At first, one might assume that sports analytics do not raise data protection issues because they will be based on information about players' performance in public games. But sports analytics also go beyond these kinds of data sets. Thus, AC Milan, a leading soccer team in Europe, uses "predictive models to prevent player injuries by analyzing physiological, orthopedic, and psychological data from a variety of sources."[25] Sports analytics of this type do raise data protection concerns.

This discussion suggests that a firm line cannot be established in advance between those analytics that require data protection safeguards and those that do not. Rather, an ongoing contextual approach is needed. The necessary kind of scrutiny will concern the actual program and the context of data use. Such an analysis will focus on the risks that analytics pose to privacy and the kinds of responsible processes that should accompany their use. These processes should also be tailored to the different stages in which analytics take place. This white paper discusses these stages below at Part III.E.

As a final point, this contextual analysis does not mean that all collection and use of "information" under the sun implicates data protection. As an example, some kinds of aggregate information involve pools that are large enough to be viewed, at the end of the day, as purely statistical and thus, as raising scant privacy risks as a functional matter. An example would be an organization analyzing high-level information about the population of the United States, China, and Japan, and their relative access to telecommunications. Other kinds of aggregate data are used in the development of flu trend information and machine-derived translation software based on the analysis of large databases of previous translated documents.[26] Companies may also choose to store their databases as de-identified information in order to reduce the privacy and security risks. In all these examples, a risk analysis should be used to determine the kinds of safeguards that are appropriate.

[24] Analytics is also affecting how the media reports on sports, with some articles focusing on how a team is performing in relation to its payroll. *See, e.g.*, Matthew Futterman, *The Year Money Didn't Matter*, Wall St. J., Sept. 17, 2010, at W8 (describing that in the 2010 baseball season, the correlation between payrolls and wins is not statistically significant).
[25] Davenport & Harris, Competing on Analytics, at 20.
[26] See below at Part III.D.

## B. Changes in the Environment for Data Protection

Change has always been a constant in data protection law. This area of law was first established in the 1970s as a response to centralized databanks, which the government at that time largely controlled. Over the last three decades, the development of the PC and then of the Internet raised new privacy concerns to which the law responded. As noted above, the contemporary landscape is marked by a "data deluge," an increasing capacity of IT, and a rise in analytics use. Leading companies and regulators are responding to these changes with a willingness to reassess past approaches. Within the corporate world, there has been a significant professionalization of privacy management. Within the ranks of the regulators, moreover, there has been an important ongoing discussion about the "modernization" of data protection law and the usefulness of an "accountability" principle.

Regarding professionalization, companies during the first decade of the 21st century have invested significant resources in data protection programs. There has been a professionalization of corporate data protection, which in turn, has been accompanied by a greater investment of business resources in this area.[27] Many corporations have come to view information privacy as important and worthy of ongoing attention. These organizations have increased investment in this area and also have assigned the development and management of privacy policies to specific employees. These employees, in turn, increasingly self-identify into categories such as Chief Privacy Officers or Chief Information Security Officers. This professionalization has provided an important precondition for a successful reliance on an accountability principle, which I will discuss shortly.

Regarding modernization of data protection law, an important and ongoing discussion has taken place over the last decade regarding the necessity of reforming or updating this area of law for the 21st century. As an example, the Federal Minister of the Interior in Germany in 2001 published an independent expert opinion, Modernization of Data Protection Law, by Alexander Roßnagel, Andreas Pfitzmann, and Hansjurgen Garstka.[28] Among the authors' many valuable points is the observation that it is now obsolete to consider the exchange or transfer of personal data as a somehow exceptional circumstance.[29] Roßnagel and his co-authors also call for new conceptual approaches to data protection law because personal information is now processed in international data networks by many participants, and frequently without possibilities for centralized governmental control.[30]

This expert opinion identified a series of core tasks that should be undertaken as part of a modernization of data protection law, including the possible use of self-regulation by industry once the law establishes a legal and regulatory framework for this behavior. They refer to this approach as "regulated self-regulation."[31] More recently, Ira Rubinstein has similarly developed a nuanced model for "co-regulation" by government and the private sector.[32] Rubinstein also calls on Congress to structure new safe harbor programs as part of the enactment of wide-reaching consumer privacy legislation.[33]

Another important part of the discussion of data protection reform has concerned the twin concepts of "accountability" and "proportionality."[34] In July 2010, the Article 29 Working Group acknowledged the need for a "statutory accountability principle" as a way to help move data protection from theory to practice.[35] Accountability

[27] *See generally* Kenneth A. Bamberger & Deirdre Mulligan, *Privacy on the Books and on the Ground*, 63 Stan. L. Rev. __ (forthcoming 2011).
[28] Alexander Roßnagel, Andreas Pfitzmann, & Hansjurgen Garstka, Modernisierung des Datenschutzrechts (2001).
[29] *Id*. at 22.
[30] They observe, "In the Internet, there is no control at the border." *Id*. at 26.
[31] *Id*. at 158–59.
[32] Ira S. Rubinstein, *Privacy and Regulatory Innovation: Moving Behind Voluntary Codes*, 6 I/S: J. L. & Pol'y Info. Soc'y __ (forthcoming 2011).
[33] *Id*.
[34] Neil Robinson et al., Review of E.U. Data Protection Directive: Summary 10 (RAND Europe May 2009).
[35] Article 29 Data Protection Working Party, Opinion 3/2010 on the Principle of Accountability, at 3.

concepts involve data controllers implementing "appropriate and effective measures" that are "scalable"—that is, proportionate to the risks involved.[36] As the Working Party has discussed this concept, "the types of measures should be coherent with the risks represented by the data processing and the nature of data."[37] Greater dangers require more significant safeguards. The goal of the Article 29 Working Group is to re-cast the Data Protection Directive to include clearer descriptions of high level outcomes and "effective enforcement measures to ensure accountability."[38]

As we will see later, the APEC Privacy Framework also supports use of accountability and proportionality as important principles in data protection. There are also other examples of an emerging international consensus regarding the value of an accountability principle.[39] Drawing on this work, this white paper will propose ways to use accountability and proportionality in an ethical use of analytics.

## C. The Ethics of Analytics and "Good Apples"

This white paper seeks to identify how companies can engage in an ethical use of analytics. This issue can be thought of as one concerning the "right action." Private-sector organizations view analytics as a powerful tool for reaching beneficial outcomes. At the same time, many companies want to be good corporate citizens and be confident that they use analytics ethically. How does one approach the use of analytics in a fashion that yields a right rather than a wrong approach? What are the principles that companies should rely on in making choices about the use of analytics?

Ethical standards matter. Firms have already sought to develop ethical labor standards and to be responsible stewards of the environment. Evidence suggests that there can be a virtuous circle from corporate agreement on ethical standards. Moreover, social science research has indicated the significant extent to which many firms seek to be "good apples."[40] It has identified numerous reasons for this result, including corporate concern about maintaining a good reputation.[41] Moreover, a company that sets a single, high baseline for its operations throughout the world can gain a considerable benefit in managing its global operations. As Samuel Palmisano, the President, CEO, and Chairman of IBM, has noted, "A company's standards of governance, transparency, privacy, security, and quality need to maintained even when its products and operations are handled by a dozen

[36] *Id*. at 9, 19.
[37] *Id*. at 19.
[38] Article 29 Data Protection Working Party, Opinion 3/2010 on the Principle of Accountability, at 9.
[39] Other examples of the accountability principle can be found. In Canada, the Office of the Privacy Commissioner has developed an international guideline that explains how the Canadian data privacy statute, the Personal Information Protection and Electronic Documents Act (PIPEDA), applies to international data transfers. In the document, the Canadian privacy commissioner stresses "an organization-to-organization approach." As Principle 1 of PIPEDA states, "An organization is responsible for personal information in its possession or custody, including information that has been transferred to a third party for processing." Office of the Privacy Commissioner of Canada, Guidelines for Processing Personal Data Across Borders (Jan. 2009), *available at* http://www.priv.gc.ca/information/guide/2009/gl_dab_090127_e.cfm.

As a second example, a review of the E.U. Data Protection Directive by RAND Europe, as commissioned by the U.K. Information Commissioner's Office, called for a re-casting of the Directive with, among other principles, an accountability concept. The RAND study wished to see clearer descriptions of high-level outcomes and implementation of "effective enforcement measures to ensure accountability." Neil Robinson et al., Review of E.U. Data Protection Directive: Summary 10 (RAND Europe May 2009).

Finally, the Galway project seeks to develop "commonly-accepted elements of privacy accountability." This effort is led by a group of international experts, facilitated by the Irish Data Protection Commissioner, and assisted by the Hunton & Williams Centre for Information Policy Leadership, which serves as secretariat to the project. Centre for Information Policy Leadership, Data Protection Accountability: The Essential Elements, A Document for Discussion (Oct. 2009), *available at* http://www.huntonfiles.com/files/webupload/CIPL_Galway_Accountability_Paper.pdf.
[40] *See, e.g.*, Neil Cunningham, Robert A. Kagan & Dorothy Thorton, Shades of Green: Business, Regulation, and Environment (2003) (explaining why certain pulp manufacturing mills go substantially beyond compliance with environmental regulatory standards).
[41] Dorothy Thorton, Neil A. Gunningham, & Robert A. Kagan, *General Deterrence and Corporate Environmental Behavior*, 27 Law & Pol'y 262, 263–67 (2005).

organizations in as many countries."[42] Even as enterprises are increasingly built around distributed business strategies, as Palmisano notes, the ultimate goal is in winning and keeping the trust of customers and business partners. Evidence shows as well that legal enforcement matters in building a virtuous circle. In particular, a fear of some level of regulatory enforcement will promote compliance. In the "implicit general deterrence theory" of Thornton, Gunningham & Kagan, for example, firms formulate compliance strategies based on a combination of their view of legal and non-legal factors.[43] Thus, "social, economic, and normative pressures" will have a significant impact on a company's behavior. Moreover, firms also pay attention to "large and dramatic formal sanctions" placed on similar organizations.[44]

[42] Samuel J. Palmisano, *The Globally Integrated Enterprise*, 85 Foreign Aff. 127, 134 (May/June 2006).
[43] Thorton, Gunningham, & Kagan, *General Deterrence*, at 265–67.
[44] *Id*. at 280.

## III. ANALYTICS IN ACTION

This white paper will discuss four areas of analytics in action to provide background on their use. These areas are: (1) multichannel marketing, (2) fraud prevention, (3) health care research, and (4) products for direct use by individuals: financial software, flu trends, and translation software.

### A. Multichannel Marketing

The use of analytics in multichannel marketing allows customized campaigns to understand the needs of consumers and align products and services with them. These can take place on a national or international basis. As an example of a domestic use of marketing analytics, Acxiom developed a "My Circle" campaign for Alltel Wireless.[45] The goal was to increase growth in new customers and reduce customer turnover, or churn. Its approach was to deliver a set of products to Alltel that included targeted segmentation of prospective customers and a "daily customer lifecycle management engine" that delivered precisely tailored marketing messages at the appropriate time in the customer lifecycle.[46] Finally, Acxiom provided advanced analytics to deepen Alltel's understanding of its prospects and customers and to provide detailed reporting on the effectiveness of its marketing campaigns.

Analytics are also used in serving ads through ad networks. These entities draw on extensive databases of people's online behavior. While ad networks generally do not have people's name, they collect information about a wide variety of interests and behavior, such as one's favorite restaurant, recent purchases, favorite movies, and TV shows. The tracking technology used by these companies include traditional cookies, flash cookies, and web beacons. As noted above, many companies involved in this business do not track names, but use these software devices to build personal profiles. The personal profiles can be primarily associated with a single code placed on an individual's computer. In one case, for example, the file consisted of this alphanumerical string: 4c812db292272995e5416a323e79bd37. These codes are used to decide which ads people see as well as the kinds of products that are offered to them. Thus, Capital One Financial Corp. draws on [x+1], an ad network, to decide instantaneously which credit cards to show first-time visitors to its website.[47]

Ad networks raise some of the most controversial issues about analytics use today. For some, this kind of tracking is changing the status quo of Internet privacy. As the Wall Street Journal has stated, "the analytical skill of data handlers . . . is transforming the Internet into a place where people are becoming anonymous in name only."[48] While ad networks may not know people's name, identification of individuals is theoretically possible.[49] This result follows because enough pieces of information linked to a single individual, even in the absence of a name or Social Security Number, can permit identification. Another objection to ad networks using analytics is the potential for steering choice. For Nicholas Carr, the "evil twin" of the "personalization" that analytics can provide is "manipulation" of future behavior.[50]

---

[45] Press Release, Acxiom Corporation, Acxiom's Direct Marketing Database Solution Helps Alltel Wireless Achieve "My Circle" Success (Oct. 13, 2008), *available at* http://www.acxiom.com/news/press_releases/2008/Pages/AcxiomsDirectMarketingDatabaseSolutionHelpsAlltelWirelessAchieveMyCircleSuccess.aspx.

[46] *Id*.

[47] Emily Steel & Julia Angwin, *On the Web's Cutting Edge, Anonymity in Name Only*, Wall St. J., Aug. 4, 2010, *available at* http://online.wsj.com/article/SB10001424052748703294904575385532109190198.html.

[48] *Id*. For another report in this series on "What They Know," see Julia Angwin, *The Web's New Gold Mine: Your Secrets*, Wall St. J., July 30, 2010, *available at* http://online.wsj.com/article/SB10001424052748703940904575395073512989404.html.

[49] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. Rev. 1701 (2010).

[50] Nicholas Carr, *Tracking is an Assault on Liberty, With Real Dangers*, Wall St. J., Aug. 7, 2010, *available at* http://online.wsj.com/article/SB10001424052748703748904575411682714389888.html. Others have defended ad networks. Jim Harper suggests that web users think about what they are getting in return for the use of tracking technology. He argues that advertising supports free content. He writes, "If Web users supply less information to the Web, the Web will supply less information to them." Jim Harper, *It's Modern Trade: Web Users Get as Much as They Give*, Wall St. J., Aug. 7, 2010, available at http://online.wsj.com/article/SB10001424052748703748904575411530096840958.html.

## B. Preventing Fraud and Protecting Data Security

Analytics are important for fraud prevention. Here, valuable examples involve MasterCard and ID Analytics. MasterCard has access to large amounts of data due to its role in processing payments between the banks of merchants and the banks of purchasers that use its charge and debit cards. This information is found across different card issuers and merchants. This access allows MasterCard to spot early indications of suspicious activity in consumer complaints, charge backs, and other areas. The company's access to information has also allowed it to become aware of fraudulent activity quickly and to take strong steps to combat it.

As a further example, ID Analytics offers a product called the "ID Score."[51] This analytics product relies on data within the company's "ID Network," which is a "real-time, cross-industry compilation of identity information."[52] The ID Network has 700 billion aggregated attributes and 2.6 million fraud reports. Through this information, ID Analytics devises an "ID Score" to support "informed customer management decisions."[53] This product permits increased accuracy in detecting identity fraud detection in real time. ID Score has been used, for example, by a top-tier financial services company to isolate nearly half of the fraud-related losses within the riskiest two percent of its customer population.[54] The product also permitted one of the largest U.S. credit card issuers to determine that two-thirds of its identity frauds fell in the category of "synthetic identities."[55] This term refers to the combining of real and fake information, including stolen Social Security Numbers, to create a new, false identity.

Analytics also play an important role in data security. In 2007, Ted Janger and I argued that a multi-institutional response was necessary to combat data security breaches.[56] One of the most important requirements of such a response was the sharing of information about security attacks among different entities to minimize harm and to increase the relevant knowledge among private organizations, governmental entities, and the public. Elements of the kind of coordinated response that we propose are now beginning to emerge. Companies in the private sector now offer services that draw on information from multiple organizations to spot data anomalies that can identify malicious activities.

Here is an area where traditional concepts of data minimization may lead to weaker rather than greater security.[57] The established idea of data minimization in the legal literature is that privacy and security would necessarily be heightened if organizations simply collected less information. As an example, James B. Rule in Privacy in Peril (2007) makes this passionate plea, "Serious privacy advocates could do well to embrace the . . . slogan: 'Less!' They need to advocate urgent programs to identify, and to help implement, ways of dealing with people that simply require less personal data from the start."[58] The examples of ID Analytics and the Schwartz-Janger

---

In a recent development, the Article 29 Working Group has recently issued an opinion concerning online behavioral advertising. Article 29 Data Protection Working Party, Opinion 2/2010 on Online Behavioural Advertising (June 22, 2010). It finds that advertising network providers are required under the ePrivacy Directive to obtain the informed consent of data subjects before placing cookies or similar devices on computers. The Working Group also calls for ad network providers to limit the time for which such consent will be valid, offer the possibility for easy revocation of it, and create visible tools to be displayed when monitoring takes place.

[51] *ID Score Datasheet*, ID Analytics, http://www.idanalytics.com/assets/pdf/ID_Analytics_ID_Score_Datasheet_Mar09.pdf (last visited Oct. 5, 2010).

[52] *ID Network Datasheet*, ID Analytics, http://www.idanalytics.com/assets/pdf/ID_Network_Attributes_Datasheet.pdf (last visited Oct. 5, 2010).

[53] *ID Score Datasheet*, ID Analytics, at 1.

[54] News Release, ID Analytics, ID Score Account Takeover Accurately Pinpoints Account Takeover at Any Point in the Customer Lifecycle (May 11, 2010), *available at* http://www.idanalytics.com/news-and-events/news-releases/2010/5-11-2010.php.

[55] *ID Score Datasheet*, ID Analytics, at 2.

[56] Paul M. Schwartz & Edward J. Janger, *Notification of Data Security Breaches*, 105 Mich. L. Rev. 913 (2007).

[57] Thus, the White House is proposing as part of its approach to enhancing online security and privacy the following "high priority action" for stakeholders in the Identity Ecosystem: "Minimize data aggregation and linkages across transactions in the Identity Ecosystem." National Strategy for Trusted Identities in Cyberspace 27 (Draft June 25, 2010) (Recommendation A 4).

[58] James B. Rule, Privacy in Peril 193 (2007).

proposal for a coordinated response entity suggest that certain kinds of barriers to data sharing and data collection might mean less rather than more data security.[59]

## C. Health Care Research

The use of analytics in health care research has created great social benefits. The National Academy of Sciences has drawn a distinction between clinical trials, a traditional form of health care research, and the new "information based" forms of inquiry.[60] In clinical trials, patients volunteer to participate in specific studies that test new medical interventions. In informational research, however, there is "the analysis of data and biological samples that were initially collected for diagnostic, treatment, or billing purposes, or that were collected as part of other research projects."[61] This technique is widely used today in sub-categories of research including epidemiology, health care services, and public health services. It also raises issues about the permissibility of secondary use of information.

The National Academy of Sciences has noted that information-based forms of health research "have led to significant discoveries, the development of new therapies, and a remarkable improvement in health care and public health." For example, through analysis of the records of a cohort of 9,000 breast cancer patients, scientists were able to identify the HER-2 oncogene.[62] Scientists then developed a targeted therapy, Heceptin, that is effective with women with HER-2 breast cancer. In another major research effort, one that started in 2003, universities, the drug and medical-imaging industries, and non-profit groups joined in a collaborative effort to find biological markers that show the progression of Alzheimer's disease in the human brain. The key element of the project was the commitment of participants to share all the data from it with the public. As the *New York Times* summarized, "The key to the Alzheimer's project was an agreement as ambitious as its goal: . . . to share all the data, making every single finding public immediately, available to anyone with a computer anywhere in the world."[63] There have already been more than 3,200 downloads of the entire data set, and almost a million downloads of the database that contains images from brain scans.[64] As use of electronic health information increases, the ability to carry out analytics on medical data will increase.

There have also been important results from medical database research that have improved the health of children. These include the discovery that supplementing folic acid during pregnancy can prevent neural tube birth defects and the identification of the negative effects of intrauterine DES exposure. In a more recent study that drew on database analysis, Flaura Winston of Children's Hospital of Pennsylvania and other researchers drew on "child-focused crash surveillance information" reported to the State Farm Insurance Companies in 15 states and the District of Columbia and then shared with the Partners for Child Passenger Safety.[65] This research was able to document the significant negative consequences when children are prematurely taken out of child safety seats or booster seats. It also found that only 25 percent of children between three and seven years of age had worn appropriate restraints in car crashes, and that children who only wore seat belts were at a 3.5-fold increase of risk of serious injury from a crash.

Finally, the study also showed that an age-appropriate child restraint, whether a child safety seat for children younger than age 4 or booster seats for children age 4 and older, could reduce the amount of head injuries in a crash. A rapid change in behavior of the target group also followed this report; parents in the U.S. markedly

[59] Regarding the need for data sharing for cybersecurity research, see Aaron J. Burstein, *Amending the ECPA to Enable a Culture of Cybersecurity Research*, 22 Harv. J. L & Tech. 167 (2008).
[60] Institute of Medicine of the National Academies, Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research 112 (Sharyl J. Nass et al., eds., 2009).
[61] *Id*.
[62] *Id*. at 114.
[63] Gina Kolata, *Sharing of Data Leads to Progress on Alzheimers*, N.Y. Times, Aug. 12, 2010, http://www.nytimes.com/2010/08/13/health/research/13alzheimer.html?fta=y.
[64] *Id*. The data itself is posted at Alzheimer's Disease Neuroimaging Initiative, http://adni.loni.ucla.edu/ (last visited October 28, 2010).
[65] Flaura K. Winston et al., *The Danger of Premature Graduation to Seat Belts for Young Children*, 105 Pediatrics 1179 (2000).

increased their adoption of booster seats for children who had outgrown car seats. As the National Academy of Sciences concludes, “Appropriate restraint by children in this age group has doubled, and child fatality from crashes is at its lowest level ever.”[66]

## D. Products for Direct Use by Individuals: Financial Software, Flu Trends, Translation Software

In some cases, a consumer can make direct use of analytics in a product. Initial examples of this category involve Mint, a leading website for management of personal finances, and Amazon, the online retailer. Further examples include the identification of flu trends and translation software.

Mint draws on analytics to analyze a customer’s data and aggregate financial information to provide targeted recommendations in real time.[67] Mint will automatically show how much a user spends in given categories, such as restaurants or gas stations, and compare spending habits with those of other users. This comparison allows a user to identify an area where she may wish to spend less. Mint also compares a user’s bank accounts, credit cards, CD, and brokerage accounts with other services, and makes recommendations for other products and companies.

As for Amazon, it makes product recommendations to customers based on analytics that examine a customer’s browsing on its site, her previous purchases, and those of other customers who looked at or purchased similar items. It explains, “We determine your interests by examining the items you’ve purchased, items you’ve told us you own, and items you’ve rated. We then compare your activity on our site with that of other customers. Using this comparison, we are able to recommend other items that may interest you.”[68] Amazon also provides tools that allow a person to manage her browsing history and shape recommendations. These include an option to delete all items and to turn off the browsing history feature. It also has an “Improving Your Recommendations” feature that explains “Why was I recommended this?” and allows revisions of ratings and exclusions of purchases.[69]

Analytics are also used in creating other kinds of new products and services. Google Flu Trends is a free service that furthers early detection of influenza epidemics throughout the world.[70] Epidemics of seasonal influenza are a major public health issue. They cause between 250,000 and 500,000 deaths worldwide annually as well as tens of millions of respiratory illnesses.[71] There is also growing concern about the possibility of a future pandemic with millions of possible fatalities worldwide if a new strain of influenza virus emerges. Scientists at Google and the Centers for Disease Control and Prevention have developed a method of analyzing large numbers of Google search queries to track influenza-like illnesses in different parts of the world. The technique monitors health-seeking behavior, specifically the online web search queries that millions of individuals submit to the Google search engine each day.[72]

Analytics involving large amounts of aggregate information have also played an important role in devising other useful products. As a final example, Google Translate is based on the concept of statistical machine translation. As the relevant FAQ explains, “[W]e feed the computer billions of words of text, both monolingual text in the

[66] *Id*. at 115. As a final example, a study of electronic health records in the use of clinical alerts to improve influenza vaccine delivery in pediatric primary care found that these alerts were associated with improvement, albeit modest ones, in increasing vaccination rates. Alexander Fiks et al., *Impact of Electronic Health Record-Based Alerts on Influenza Vaccination for Children with Asthma*, 124 Pediatrics 159 (2009).
[67] *Our Product*, Mint, http://www.mint.com/ (last visited Oct. 6, 2010).
[68] *Help: Recommendations*, Amazon, http://www.amazon.com/gp/help/customer/display.html/ref=hp_left_cn?ie=UTF8&nodeId=13316081 (last visited Oct. 6, 2010).
[69] *Id*.
[70] Jeremy Ginsberg et al., *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 Nature 1012 (Feb. 19, 2009).
[71] Fact Sheet N°211, World Health Organization, Influenza (Seasonal) (Apr. 2009), *available at* http://www.who.int/mediacentre/factsheets/fs211/en/.
[72] Ginsberg, *Detecting Influenza Epidemics*, at 1014.

target language, and aligned text consisting of examples of human translations between the languages. We then apply statistical learning techniques to build a translation model."[73] Much of the "aligned text" consists of official translations from the United Nations and European Parliament as well as other sources.[74] Other software-driven translation products are offered by Microsoft and IBM. Beyond these examples of analytics in action, there are also distinct steps within a project that takes a data-driven approach. This White Paper now considers the steps that typically occur when a firm uses analytics.

## E. 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, and, as I will argue below, Fair Information Practices (FIPs) should be tailored to make an effective contribution to promoting privacy at all four stages.

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 E.U. Data Protection Directive and other E.U. 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 E.U. Data Protection Directive has an expansive definition of the concept of the "processing of personal data."[75] 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."[76]

This approach might also mean finding a way to get a proprietary advantage by improving collection of information in one's own data operations, creating new kinds of relations with customers that will generate new kinds of information, or using commercially available data. The search for unique information encourages each company to seek new avenues for data collection. As noted above, translation products have drawn on large data sets of publicly available, multiple translations of the same documents from the United Nations and European Parliament. In search of an additional data set, Google has also drawn on information from its book-scanning project.

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

[73] Franz Och, *Statistical Machine Translation Live*, Google Research Blog (Apr. 28, 2006 3:40 PM). http://googleresearch.blogspot.com/2006/04/statistical-machine-translation-live.html; *see also Inside Google Translate, How Does it Work?*, Google, http://translate.google.com/about/intl/en_ALL/ (last visited Oct. 6, 2010). Google Translate Services can handle 57 languages.
[74] For a discussion, see Miguel Helft, *Google's Computing Power Refines Translation Tool*, N.Y. Times, Mar. 8, 2010, *available at* http://www.nytimes.com/2010/03/09/technology/09translate.html?_r=1&pagewanted=print.
[75] *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].
[76] Davenport, Harris & Morison, Analytics at Work, at 25.

organization."[77] As an example mentioned above, the Partners for Child Passenger Safety established a database of claims from an insurance company about accidents involving children in different jurisdictions.[78] 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. Thus, MasterCard obtains access to detailed purchase information across card issuers. This data allows it to assist banks in co-branding or creating interesting offers; the data also helps MasterCard understand why a given cardholder might be using a card less.

As a final example, Harrah's, a private gambling corporation, provides customers with loyalty cards that follow their behavior in real time. At the aggregate level, marketing and operations at Harrah's look at data "to optimize yield, set prices for slots and hotel rooms, and design the optimal traffic flow through the casinos."[79] But the casino operator also sends messages to customers at different times based on the loyalty card data. It might decide that a person is losing "too much money too fast." A message will then be sent to the person, such as "Looks like you're having a tough day at the slots. It might be a good time to visit the buffet. Here's a $20 coupon you can use in the next hour."[80]

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."[81] 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 renewal."[82] They recognize the need for "carefully monitoring outcomes and external conditions to see whether assumptions need to be modified."[83] Amazon is one of the companies that has actually 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.

[77] *Id*. at 29.
[78] Flaura K. Winston et al., *The Danger of Premature Graduation to Seat Belts for Young Children*, 105 Pediatrics 1179, 1179 (2000).
[79] Davenport & Harris, Competing on Analytics, at 85.
[80] *Id*. at 86.
[81] Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 Tex. L. Rev. 669, 676 (2010).
[82] Davenport & Harris, Competing on Analytics, at 109.
[83] *Id*. at 129.

# IV. THE FIT WITH FAIR INFORMATION PRACTICES (FIPS)

Analytics raise complex questions for data protection law built around traditional fair information practices. In this Part, I wish to identify two areas in which there is tension with some aspects of FIPs. The white paper will discuss FIPs as expressed both in the OECD Guidelines (1980), the EU Data Protection Directive (1995), and the Asian-Pacific Economic Cooperation (APEC) Privacy Framework (2004).

The areas of tension with analytics concern the principles of (1) automated individual decisions and (2) the related concepts of purpose specification and use limitation. As we will see, the OECD Guidelines and APEC Framework differ from the Data Protection Directive in their approach to the concept of automated decision-making. As for purpose specification and use limitation, the OECD Guidelines, Data Protection Directive, and APEC Framework have a generally consistent approach to these principles. At a minimum, rules about the collection and processing of personal information do reflect, at least for certain questions and in certain contexts, different legal, social, and cultural values. The challenge for private-sector organizations operating in different countries and in an age of international data flows will be to harmonize their global operations around a single set of ethical standards.

## A. Automated Individual Decisions

The first area for consideration concerns automated decision-making. Here, there are differences between the OECD Guidelines and APEC Framework, on the one hand, and the Data Protection Directive, on the other. In a nutshell, the OECD Guidelines and APEC Framework do not treat automated decision-making as distinct from overall issues of data protection while the Data Protection Guidelines provide special protections for it.

The Explanatory Memorandum to the OECD Guidelines explains its integrated structure for data protection and its decision to treat automated decision-making as only part of its overall structure and not as a category apart.[84] The OECD Guidelines declare:

Above all, the principles for the protection of privacy and individual liberties expressed in the Guidelines are valid for the processing of data in general, irrespective of the particular technology employed. The Guidelines therefore apply to personal data in general or, more precisely, to personal data which, because of the manner in which they are processed, or because of their nature or context, pose a danger to privacy and individual liberties.[85]

As the Memorandum also explains, the difficulty of making a “clear distinction” between automatic and non-automatic processing contributes to its choice not to regulate automatic data processing separately.

From today’s perspective, the OECD Guidelines from 1980 are an important forerunner for the contemporary interest in accountability and proportionality.[86] Rather than deciding to focus only on the threat to privacy from a certain technology, namely, automated decision-making, the Guidelines took a contextual approach to assessing privacy risks. To return to the language quoted above, the OECD Guidelines sweep in all “personal data which,

[84] It sets the stage by noting, “The 1970s may be described as a period of intensified investigative and legislative activities concerning the protection of privacy with respect to the collection and use of personal data.” Organization for Economic Co-operation and Development [OECD], *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, Explanatory Memorandum, ¶ 1 (Sept. 23, 1980). The Explanatory Memorandum to the Guidelines quickly turns to address the “legal problems of automatic data processing.” *Id.*, ¶ 3. It notes the “widespread concern” regarding “privacy and individual liberties” regarding such use of computers and also discusses the reasons for it. *Id*. The Explanatory Memorandum’s analysis, expressed in 1980, appears quite prescient as it touches both on the “the ubiquitous use of computers” as well the creation of “complex national and international data networks” through the use of computers and telecommunications. *Id.*.

[85] *Id.*, ¶ 37.

[86] The OECD Guidelines also explicitly decline to define the concept of “automatic data processing.” *Id.*, ¶ 36. The Explanatory Memorandum explains that the choice follows because it was so difficult “to make a clear distinction between the automatic and non-automatic handling of data.” *Id.*, ¶ 35. It added, “There are, for instance, ‘mixed’ data processing systems, and there are stages in the processing of data which may or may not lead to automatic treatment.” *Id*.

because of the manner in which they are processed, or because of their nature or context, pose a danger to privacy and individual liberties."[87]

Like the OECD Guidelines in 1980, the APEC Privacy Framework in 2004 takes a contextual approach. It also introduces important concepts concerning accountability and proportionality. The APEC Framework states that organizations that process information should be "accountable for complying with measures that give effect" to its principles.[88] Like the OECD Guidelines, it does not single out automated processing for special protections. Rather, the APEC Framework's attention is on preventing harm from "misuse of personal information."[89] It requires, in a fashion similar to the OECD's contextual approach, that "remedial measures should be proportionate to the likelihood and severity of the harm threatened by the collection, use and transfer of personal information."[90]

The APEC Framework also extends this principle of proportionality by calling for "accountability." In its language, "A personal information controller should be accountable for complying with measures that give effect" to its principles.[91] As I have noted, other international organizations have adopted the ideas of accountability and proportionality and, as will be developed below, these concepts are extremely helpful as touchstones for developing an ethical use of analytics.

In contrast to the contextual approach to the question of automated decisions in the OECD Guidelines and APEC Framework, the Data Protection Directive singles out automated processing for special safeguards. As the Directive states, "Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him . . . based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc."[92] Exceptions to this rule are permitted only if the automated decision is taken pursuant to entering into a contract or authorized by a law "which also lays down measures to safeguard the data subject's legitimate interests."[93] The path of special treatment is marked by special rules concerning the right of access to "the logic involved in any automatic processing of data concerning" an individual and a required notification of supervisory authorities of automatic data processing.[94]

## B. Purpose Specification and Use Limitations

The OECD Guidelines, EU Data Protection Directive, and APEC Privacy Framework treat the concepts of purpose specification and use limitations in a generally consistent fashion. Here, there is an interesting difference with the general approach in the United States, which starts from a perspective that favors the free flow of information. These different starting points should not be overemphasized, however, as the law in the United States protects information privacy in many circumstances, and is willing to place limits on the free flow of information to do so.

[87] *Id*., ¶ 37.
[88] Asia-Pacific Economic Cooperation [APEC], *APEC Privacy Framework*, Part III, Principle IX, ¶ 26 (2005).
[89] *Id*., Part III, Principle I, ¶ 14.
[90] *Id*.
[91] *Id*., Part III, Principle IX, ¶ 26.
[92] 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, Section VII, art. 15(1), 1995 O.J. (L 281) [hereinafter Data Protection Directive].
[93] Id., Section VII, art. 15(2).
[94] Concerning the right of access, the Directive grants every data subject the right to obtain a wide set of information from the data controller, including information as to the purposes of the processing and "knowledge of the logic involved in any automatic processing of data concerning him." *Id*., Section V, art. 12(a). As for the required notification of supervisory authorities of automatic data processing, the Directive states that controllers are to notify supervisory authorities before "carrying out any wholly or partly automatic operation." *Id*., Section IX, art. 18(1). Exceptions are permitted to be made by Member States for a number of situations including (1) the unlikelihood of a violation of the interests of the data subjects, and/or (2) the appointment of a data protection official at the data controller. Id., Section IX, art. 18(2).

Nonetheless, the different starting points can lead to diverse outcomes, legal and otherwise, for programs of personal information processing.[95]

We begin this exploration of purpose specification and use limitations with the OECD Guidelines. This document requires the specification of "the purposes for which personal data are collected . . . not later than at the time of data collection."[96] It also calls for a compatibility limitation on any subsequent use. Compatibility means the use of information should be "limited to the fulfillment of the [initial] purposes for collection or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose."[97]

The Data Protection Directive also has a specific category concerning purpose specification. According to it, personal data must be "collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes."[98] In a similar fashion, the APEC Privacy Framework states, "The collection of personal information should be limited to information that is relevant to the purposes of collection."[99]

In both the OECD Guidelines and the EU Data Protection Guidelines, moreover, there are rules that create use limitations. As the OECD Guidelines state, personal data is to be collected only "by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject."[100] It also permits the use of personal data only for purposes originally specified, "with the consent of the data subject," or "by the authority of law."[101]

The Data Protection Directive approaches use limitations through the requirement of limited categories for making data processing legitimate. According to the Directive, personal data may be processed only pursuant to a limited group of categories, such as unambiguous consent, fulfillment of consent, compliance with a legal obligation, protection of a vital interest of the data subject, a task in the public interest, or a legitimate interest of the data controller.[102] In addition, in the case of processing in the public interest or for a legitimate interest of the data controller, the data subject is to be given a right to object to the processing.[103]

The APEC Privacy Framework also contains limited categories for permitting data processing. It allows the use of collected personal information "only to fulfill the purposes of collection and other compatible or related purposes,"[104] Its exceptions include consent of the individual and provision of a service or product that the individual has requested, pursuant to a law.

The reliance on these limited categories in the Directive and APEC Privacy Framework presents a notable contrast with the approach in the United States. At the risk of oversimplification, the legal approach in the United States generally permits the use of personal information unless a law prohibits it. This level of legal regulation in the United States is due, in part, to the strong First Amendment protections for freedom of expression. In contrast, the Data Protection Directive and APEC Privacy Framework prohibit the collection and processing of personal data unless a law explicitly permits it, or another limited set of permissive categories is involved. Thus, the starting

[95] For an exploration of these differences in legal outcomes for tort privacy in Germany and the United States, see Paul M. Schwartz & Karl-Nicholas Peifer, *Prosser's Privacy and the German Right of Personality*, 98 Calif. L. Rev. __ (forthcoming 2010).
[96] OECD Guidelines, Part II, Principle 9.
[97] *Id*.
[98] Data Protection Directive, Section I, art. 6(1)(b).
[99] APEC Privacy Framework, Part III, Principle III, ¶ 18.
[100] OECD Guidelines, Part II, Principle 7.
[101] *Id*., Part II, Principle 10.
[102] Data Protection Directive, Section II, art. 7.
[103] *Id*., Section VII, art. 14.
[104] APEC Privacy Framework, Part III, Principle IV, ¶ 19.

points about information collection can be different in different countries, and this diversity can lead to varying results concerning the permissibility of certain kinds of programs of personal information processing.[105]

Further, these disparities demonstrate that rules about the collection and processing of personal information can reflect different legal, social, and cultural values. The sociologist Christena Nippert-Eng has recently noted her conviction that "privacy and what is considered private or public are such culturally specific ideas and practices," and that her findings about privacy in the United States cannot simply be applied without further research of other societies.[106] Nippert-Eng also observes that her international travels have convinced her as well "just how central whatever happens in the U.S. can be for shaping conversations and practices and policies throughout the world."[107] Other scholars have explored the comparative dimensions of privacy and privacy regulation in different societies and legal systems.[108]

At the same time, certain kinds of data collection in the United States, although not prohibited by law, will nonetheless strike Americans as violative of non-legal norms of proper behavior. In public and policy discussions, the language describing such data processing will speak of inappropriate actions, creepiness, intrusiveness, or rudeness. Part of the ethical use of analytics involves consideration of these non-legal dimensions of standards of good behavior. This white paper returns to this point below.

[105] For an exploration of these differences in legal outcomes for tort privacy in Germany and the United States, see Paul M. Schwartz & Karl-Nicholas Peifer, *Prosser's Privacy and the German Right of Personality*, 98 Calif. L. Rev. __ (forthcoming 2010). This paper also finds numerous areas in which the different legal systems for tort privacy converge on similar solutions.
[106] Christena Nippert-Eng, Islands of Privacy 16–17 (2010).
[107] *Id*. at 17.
[108] *See, e.g.*, Schwartz & Peifer, *Prosser's Privacy and the German Right of Personality*; Colin J. Bennett & Charles D. Raab, The Governance of Privacy (2006); James Q. Whitman, *The Two Western Cultures of Privacy*, 113 Yale L.J. 1151 (2004).

# V. REVISITING FIPS FOR ETHICAL USE OF ANALYTICS

In Part III.E, this white paper identified four discrete stages of analytics. These are: (1) collection, (2) integration and analysis, (3) decision-making, and (4) review and revision. In this Part, we turn to the issue of the ethical use of analytics during these 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 first discuss ethical requirements in the respective step and then 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 actor. As noted above, the ethical discussion should include consideration of non-legal standards of good behavior for organizations. Such consideration is especially important because of the rapid development of IT and analytics. Early adopters and innovators in this field may develop capabilities that the law does not yet regulate. Hence, an ethical organization should also evaluate non-legal norms in assessing the impact of its use of analytics.

Moreover, trust should be an important factor in the ethical use of analytics. A company should keep in mind the value of having different stakeholders recognize it as a responsible social organization. The relevant stakeholders can include consumers, the public at large, other businesses, government, and non-governmental organizations. Trust also becomes an especially important challenge in an age of distributed services. In addition, as noted above, certain kinds of data processing may seem inappropriate, creepy, intrusive, or rude. Such activities will have a highly pernicious impact on trust in the company. In turn, the lack of trust can short circuit a company's ability to use novel kinds of data processing approaches to better serve its customers.

As an additional requirement, at each stage of the use of analytics, a company should have accountable processes that provide safeguards proportionate to the risks. As we have seen, the APEC Privacy Framework establishes an accountability concept as one of its chief privacy principles, and there are other examples of an emerging international consensus regarding the value of accountability. Accountable privacy requires processes tailored to the specific risks of analytics. An accountability analysis begins with an acknowledgment that analytics can have a negative impact on individuals. It then turns to the development of internal policies that center on rules of information management and training of personnel. It also requires that a company place responsibility for data processing operations and decisions on a set of defined individuals. Leading companies are now shifting to such a new process-oriented management approach.[109]

Another overarching requirement is for data security. The OECD Guidelines, Data Protection Directive, and APEC Guidelines all recognize this principle. For example, the OECD Guidelines provide, "Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data."[110] The Data Protection Directive requires that the controller "implement appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss."[111] The APEC Guidelines state, "Personal information controllers should protect personal information that they hold with appropriate safeguards against risks, such as loss or unauthorized access to personal information, or unauthorized destruction, use, modification, or disclosure of information or other misuses."[112] Experts in the use of analytics have also noted the importance of data security in analytics. For

[109] Kenneth A. Bamberger & Deirdre Mulligan, *Privacy on the Books and on the Ground*, 63 Stan. L. Rev. __ (forthcoming 2011); Paul M. Schwartz, Managing Global Data Privacy: A Report from the Privacy Projects (2009), *available at* http://theprivacyprojects.org/wp-content/uploads/2009/08/The-Privacy-Projects-Paul-Schwartz-Global-Data-Flows-20093.pdf.
[110] OECD Guidelines, Part II, Principle 11.
[111] Data Protection Directive, Section VIII, art. 17.
[112] APEC Privacy Framework, Part III, Principle VII, ¶ 22.

example, Davenport, Harris, and Morison note that "[h]ighly analytic organizations" collect information and then "guard it with their lives."[113]

The ideas of accountability and proportionality are of great assistance in developing workable data security principles. Data security should reasonably protect against the risks of data loss, destruction, and related harms. In other words, an organization should employ data security that is reasonable when measured against the kinds of information that are collected and processed, and the decisions that are made with it.

Finally, an organization's processing, analysis, and decision-making through analytics should respect cultural and social norms concerning the use of "sensitive" information. "Sensitivity" depends not only on an abstract category or categories in which information can be placed (health information, age, ethnicity), but on the overall context and conditions of the processing, analysis, and decision-making. Thus, a company should carry out a contextual approach to evaluating the sensitivity of a specific use of analytics. When its use of analytics involves more sensitive areas, it should employ reasonable safeguards proportionate to the risk.

To summarize, therefore, the points made thus far:

- 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. The examples about health related research above included important investigations about children. In particular, analytics involving aggregate data sets can lead to significant medical advances and public safety breakthroughs. As a result, a simple ban of 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.[114] Due to the special vulnerability of

[113] Davenport, Harris & Morison, Analytics at Work, at 34.
[114] *See* Jeff Chester & Kathryn Montgomery, Berkeley Media Studies Group, Interactive Food & Beverage Marketing: Targeting Children and Youth in the Digital Age (May 2007).

children, the use of analytics to shape contacts with them raises special ethical concerns. As one indication of the possibility for consensus around this topic, the Network Advertisement Initiative already forbids use of behavioral advertising directed toward children under 13. Psychological research indicates, however, that the period of childhood vulnerability continues well past this age. The ethical use of analytics should responsibly acknowledge the vulnerability of children and place responsible limits on the use of this technology.

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

## B. Stage One: Collection

Data Quality. The principle of "data quality" is a traditional FIP, which, nonetheless, raises potentially problematic issues if enforced at the collection stage. The OECD Guidelines, Data Protection Directive, and APEC Guidelines all acknowledge this concept. In the ethical use of personal information, as the OECD Guidelines state, companies should use information in a way "relevant to the purposes for which they should be used." Moreover, "to the extent necessary for those purposes," the information "should be accurate, complete and kept up-to-date."

Here, there is a potential tension with the focus in analytics on the collection of large amounts of information. It is through the use of analytics that companies often decide whether or not information is, in fact, relevant. This result suggests that the "data quality" FIP will be of greater use not at the collection stage, but at a later period in the use of analytics. We return to this point below.

Collection Limitation and Purpose Specification. Complex questions also arise concerning the applicability of collection limitation and purpose specification during analytics. The contemporary use of analytics favors a search for meaningful relationships among a diverse and wide set of collected information. As Kenneth Bamberger explains, "Analytic models … provide insight into the probability of specific outcomes, usually by analyzing large sets of historic data."[115] The same point about the need for gathering large quantities of information is made by Davenport, Harris, and Morison. They state, "Highly analytical organizations tend to gather a lot of information about the entities they care about most—usually customers, but sometimes employees or business partners."[116]

Even if analytics depends on extensive data, the ethical use of analytics requires that companies make appropriate choices about the collection of information. Certain information should not be collected due to legal, cultural, and social factors. The best way to assess the ethics of collection is to consider accountability and proportionality. Collection of some kinds of information may bring such risks to the company and affected individuals, including data security risks, that a company should decide to forgo the collection.

- A company should not collect certain 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

At the integration and analysis stage, companies perform analytics on their assembled data sets. Here, it is once again important that organizations carry out a risk analysis and that accountable processes be in place. For legal, cultural, and social reasons, and due to risks to the company and affected individuals, there may also be certain kinds of information that should not be integrated into the analytics process. Analysis may also reveal data that proves to be of problematic quality, and in the ethical use of personal information, companies should not rely on information once it is shown to be of insufficient data quality.

[115] Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 Tex. L. Rev. 669, 689 (2010).
[116] Davenport, Harris & Morison, Analytics at Work, at 34.

Moreover, it may be possible to carry out integration and analysis with the use of anonymous data. The use of anonymous data can help in many circumstances to reduce security concerns and threats to privacy. Even if certain kinds of anonymous data can be subject to re-personalization, a recourse to anonymization in many circumstances still leads to a net increase in security and privacy protections.

- Companies should refrain from use of information once integration and analysis shows it to be of insufficient data quality for the intended purpose.
- Companies should anonymize personal information when appropriate in their analysis of it.

## D. Stage Three: Decision-making

At the decision-making stage, the ethical use of analytics involves a number of safeguards and steps. Some use of analytics might be relatively trivial, and others more significant in terms of the impact on the individual. A lesser impact would be a company's decision whether to serve a person visiting a website an advertisement for running shoes or an advertisement for some other type of sporting goods. Other decisions, such as whether or not to offer credit, have more important consequences.

As a company makes decisions about an individual that are more significant, it should seek greater accuracy and have stronger safeguards in place. As a critical touchstone, analytical output should be reasonably accurate in proportion to the nature and significance of the underlying decision. Put differently, the concept of reasonably accurate and proportionate decision-making depends on the context of the decision made through analytics.

In addition, companies should provide reasonable notice, access to one's information, and remedies to individuals once decisions are made about them. The OECD Guidelines, Data Protection Directive, and APEC Framework all call for a range of compensatory controls for individuals. The APEC Framework provides a useful perspective in calling for "an appropriate array of remedies for privacy protection violations, which could include redress, the ability to stop a violation from continuing and other remedies."[117] It also calls for "a number of factors to be taken into account" in deciding on the necessary measures including "the importance of having a range of remedies commensurate with the extent of the actual or potential harm to individuals resulting from such violations."[118]

Thus, when risk results in harm to individuals, a company should provide appropriate mitigation and remedies. Compensatory controls should also allow individuals to express their own preferences about analytics and to reflect their own assessments about risk and other factors when appropriate. At the same time, such an interest cannot permit an individual to re-write certain information that reflects external history. It would not be appropriate, to provide just two examples, for individuals to have a right to alter freely their credit information or their educational records. The same logic applies to paper-based records.

As a recent positive trend concerning compensatory controls, companies are now offering new privacy tools to individuals. The resulting controls include the ability to opt out from tracking, and to set preferences about the kinds of information that are collected. Google, Microsoft, and Yahoo are among the companies offering such privacy tools. The Network Advertising Initiative has also developed an Opt-out Tool that allows users to examine their computer to see which NAI member companies have placed an advertising cookie file on their computer and to opt out of tailored advertisements.[119]

[117] APEC Privacy Framework, Part IV, ¶ 38.
[118] *Id*.
[119] *See Opt Out of Behavioral Advertising*, National Advertising Initiative, http://www.networkadvertising.org/managing/opt_out.asp (last visited Oct. 7, 2010).

Finally, this white paper has discussed safeguards in the E.U. Directive concerning “automated decision-making.” In different cultural and social contexts, the use of analytics to reach decisions about individuals may require additional safeguards to be considered as appropriate.

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

At the review and revision stage, companies should acknowledge that change is the only constant in the use of analytics. A business model may evolve. Databases may no longer be current. Moreover, a theoretically valid pattern revealed by analytics may not lead to the actual results foreseen. The use of analytics may also lead to unforeseen consequences that raise ethical questions.

As a consequence, the ethical use of analytics requires ongoing review of how the results track in the real world. Analytics that are not reasonably accurate should be revised. The data quality principle becomes important, therefore, at this stage. Review and revision will also help a company’s financial bottom line by making sure that they are not relying on stale theories or invalid algorithms. Finally, a company should exclude information from its analytics that is not found to be predictive. During review and revision, analysis should evaluate the extent to which information collected and integrated and analyzed was actually predictive.

- 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 overall capacities, brings with it the potential for positive and negative effects. This 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.

THE CENTRE
FOR INFORMATION
POLICY LEADERSHIP
HUNTON & WILLIAMS LLP