November 2019

ADVANCED DATA ANALYTIC PROCESSING – 2019 UPDATE

Prepared by Paula Bruening for the Information Accountability Foundation

Foreword by Martin Abrams

Policymakers around the world are engaged in discussions about whether and how to enact privacy laws that would promote economic growth, fair processing, and the benefits that flow from deployment of digital technology and the use of data. Ideally, these laws would protect individuals when they participate in digital life and foster an environment where data is used fairly, for their benefit and for the benefit of society. To be effective and forward-looking, such emerging privacy law must address the realities of advanced data analytics – powerful processing that uses vast, varied data stores to develop new insights and address problems once had resisted solutions.

In 2013 Paula Bruening, the author of the paper that follows, Meg Leta Ambrose and I released "Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance" (the "2013 Paper").[1] In it we explored how organizations might thoughtfully design governance for advanced data analytics. At that time, organizations struggled to justify the use of advanced data analytics in a regulatory environment that did not take into account the realities of how that processing occurred. The 2013 Paper suggested an approach that envisioned advanced data analytics as occurring in two phases. The first phase - *knowledge discovery* - involved robust use of data to create new insights. The second - *knowledge application* – used those

[1] "Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance," Centre for Information Policy Leadership, February 2013, https://www.huntonak.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf.

insights to make decisions that often affect individuals. We suggested that knowledge discovery was much less risky to individuals than knowledge application where decisions are actually made. Both phases give rise to data protection obligations, but the 2013 paper proposed that the governance for each phase should reflect the level of risk it raises.

The 2013 Paper, therefore, proposed that governance of knowledge discovery should provide more flexibility than that allowed for in knowledge application – to reflect the lower risk knowledge discovery poses to the individual. When applying insights in the knowledge application phase to make decisions about individuals, however, legacy privacy obligations – particularly accountability - were much more applicable.

Since the release of the 2013 Paper, technology and business process have continued to evolve at an ever-faster pace. Even recently enacted laws, such as the European Union's General Data Protection Regulation ("GDPR"), struggle with how to implement data protection measures that advance the benefits of advanced analytics while still allowing a space for seclusion, individual control, and fair processing. The purpose of the paper that follows is to focus attention on how to design governance for advanced data analytics that addresses the risks associated with the impact of data-driven knowledge application. It proposes that when designed in this way, guidance can protect individuals and still allow society to reap the benefits of data-driven knowledge discovery.

Introduction

In 2013, Abrams, Bruening and Ambrose published the 2013 Paper. The 2013 Paper highlighted the benefits and risks of advanced data analytics and the need to identify an approach to governance that addressed risks while making it possible to realize advanced data analytics' full potential. This paper revisits that work in light of developments in advanced data analytics[2] and artificial intelligence (AI), and efforts to enact privacy and data protection law, regulation and guidance. It emphasizes the expanding vision for advanced data analytics, and comments on how it can benefit society and the risks it can raise. It explains the two-phased nature of advanced data analytics - *knowledge discovery* and *knowledge application* - introduced in the 2013 Paper and how the two phases differ. By examining the evolving landscape for advanced data analytics - including developments in technology, legislation and policy – this paper highlights why it is urgent that effective and workable governance that reflects this distinction is developed and implemented. It considers how the concept of "legitimate interest," articulated first in the European Directive and currently included in the GDPR, could serve a key role in governance of advanced data analytics. Based on this discussion, this paper proposes that *knowledge creation* can be conducted in a more trusted way when organizations assess their decisions to process data for this purpose based on whether it meets the requirements of legitimate interests. It further proposes that organizations would be helped by clearly articulated guidance that takes into full account the benefits of the knowledge creation and the risks to individuals about how these assessments should be carried out.

The Findings of the 2013 Paper

The 2013 Paper highlighted the benefits and risks of advanced data analytics. It identified the need for an approach to governance that mitigates risk while making it possible to realize advanced data analytics' full potential. The 2013 Paper characterized advanced data analytics as a two-part process: *knowledge discovery,* in which data is analyzed to understand what

[2] While the 2013 Paper uses the term "big data analytics," this paper uses the more current and descriptive term "advanced data analytics" throughout.

insights and inferences it can offer and can be incorporated into an algorithm, and *knowledge application,* in which algorithms developed in the knowledge discovery phase are applied to individuals to make decisions about them. It further highlighted the challenges that compliance with current law, regulation and principles of fair information practice pose for organizations deploying advanced data analytics. These challenges included:

- how *traditional notions of consent* apply to use of advanced data analytics for knowledge discovery;
- interpretation and application of the *legitimate business purpose* analysis articulated in the GDPR to justify decisions to deploy advanced data analytics;
- the relevance and feasibility of the *purpose specification* principle to the ability to derive insights and predictions from data that are often unexpected and not foreseeable;
- the relevance of *data minimization* in light of the need for vast, varied data stores to support and realize the promise of advanced data analytics; and
- prohibitions in law against *automated individual decisions and profiling*.

In light of these challenges, the 2013 Paper set goals for advanced data analytics guidance, stressing that governance should recognize and reflect the two-phased nature of advanced data analytics, the difference between them, and the risks each does or does not raise. It suggested that the risks that arise from advanced data analytics occur predominantly in the knowledge application phase, when algorithms derived in the knowledge discovery phase are used to arrive at insights upon which organizations can base decisions about individuals. It suggested that fair information practice principles continue to serve as a trusted, commonly-recognized basis for guidance but cautioned that they must be applied in a way that serves the realities of advanced data analytics and pointed to the importance of the accountability principle to practical, effective governance. It emphasized the need for guidance about how to establish that the use of data for knowledge discovery is a legitimate business purpose when balanced with the full range of stakeholder interests. By characterizing advanced data analytics as a two-part process – knowledge discovery and knowledge application - the 2013 Paper opened a path to governance that appropriately locates and addresses the risk of advanced data analytics, especially the risk of knowledge application, and in doing so, allows more room for understanding data and for realizing its benefits.

Since 2013, the proliferation of data, and rapid advances in AI and other forms of advanced data analytics have increased attention to the benefits and risks of advanced data analytics. At the same time, calls for legislation in the U.S. and other jurisdictions have highlighted the need for workable governance. The Information Accountability Foundation (the "IAF") believes that such governance must rest on a framework that enables organizations to deploy advanced data analytics for beneficial ends, foster the trust of individuals, and establish legal certainty. It also should support the ability of organizations to be confident in their decisions to use data and not forego opportunities to address problems simply because they lack useful guidance.[3]

Advances in Advanced Data Analytics and Growing Recognition of the Risks and Benefits

The Proliferation of Data for Analytics

Data scientists rely on the proliferation of diverse and varied data stores to tap the potential of advanced data analytics. Experts predict exponential growth in the rate at which data is generated, fueled by technology, connectivity, and the creation of inferred data. The size of the digital universe is expected to double every two years at least - a 50-fold increase in rates of growth from 2010 to 2020. Human- and machine-generated data now grows 10 times faster than traditional business data, and machine data is increasing even more rapidly at 50x the growth rate.[4]

- By 2020 new data generated per second for every human being will amount to approximately 1.7 megabytes.
- By 2020, the accumulated volume of data will increase from 4.4 zettabytes to roughly 44 zetabytes (or 44 trillion GB).

[3] With the 2013 Paper as a backdrop, the IAF contributed to policymakers' efforts to address the challenges of advanced data analytics governance and arrive at workable solutions. This work generated guidance instruments founded on the two-phased nature of advanced data analytics to enable stakeholders to optimize the benefits of advanced data analytics, mitigate risk, and use data responsibly.

[4] "The Exponential Growth of Big Data," *Inside Big Data,* February 16, 2017, https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data/.

- Every minute Facebook users send approximately 31.25 million messages and watch 2.77 million videos.
- The data gathered no longer consists only of text but also includes media such as video and digital images. Every minute on YouTube alone users upload 300 hours of video suggesting the rapid rate of growth of this kind of data.
- The number of smart connected devices in the world is expected to grow to more than 50 billion – all of which will generate data that can be shared, collected and analyzed.[5]

Data available for advanced analytics will be generated not only through observation and use of technology but also through *inferred* data. Inferred data are data and insights about individuals generated based on search histories, content consumption, purchases and social media activity. Inferred data also results when an analytics system assigns a predictive value or an attribute to an individual based on what has been calculated without his or her involvement.[6]

The growth in the rate of data generation promises deeper and broader resources for advanced data analytics. Data will come from new data subjects, be generated by new activities and be more varied in kind. Data for analytics will grow in volume and variety. The sheer breadth of data amplifies its power, promising deep, unexpected and valuable insights.

The Power of Advanced Data Analytics

Advanced data analytics hold the potential for customized solutions to challenges individuals face -- such as in precision medicine and personal finance management. Advanced data analytics also promise new ways to address societal and global issues – such as disaster relief and efficient distribution of resources. Advanced data analytics are predicted to make

[5] 21 Big Data Statistics & Predictions on the Future of Big Data," *New Generation Apps,* January, 2018 https://www.newgenapps.com/blog/big-data-statistics-predictions-on-the-future-of-big-data.

[6] Furthermore, inferred data relies on the processing of data about many individuals to develop algorithms. These algorithms are then used to make predictions about an individual or cohort of individuals such as a family, neighborhood or community.

workplaces more efficient, improve schools' ability to serve students and enhance the safety of highways and the quality of life of cities. Businesses look to advanced data analytics to deepen their understanding of their customers, personalize and time their marketing efforts to make them more effective, increase productivity, and control costs.[7] When amplified by AI, advanced data analytics is predicted by 2030 to contribute to a 26% increase in local gross domestic product (GDP).[8] Advanced data analytics introduce a new way of thinking about where breakthroughs in knowledge are found and how they occur, creating pathways to discovery across nearly every research discipline.

Rapid Advances in AI

Perhaps the most powerful deployment of advanced data analytics is realized in AI. Data fuels the algorithms that form the foundation of AI; advances in AI technology and its broad adoption raise the stakes for advanced data analytics governance that promotes trust and mitigates risks to individuals.

Since 2016, experts and policymakers have turned their attention to the rapid development and adoption of AI. A recent article in MIT Technology Review highlights the rate of investment in AI startups (as compared to startups generally) and the rapid rate of technical progress in AI, citing improved accuracy of object recognition in images, measured against human performance and the accuracy of machine translations of news articles as examples. It also notes the rise of AI as a political issue, as the U.S. Congress and the UK Parliament have turned their attention to developments in the sector, reflecting a growing awareness of AI's economic and strategic importance.[9]

---

[7] "Analytics Comes of Age," McKinsey Analytics, 2018, https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Analytics%20comes%20of%20age/Analytics-comes-of-age.ashx.

[8] "PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution," PwC, 2017-2019, https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

[9] Will Knight, "Nine charts that really bring home just how fast AI is growing," MIT Technology Review, December 12, 2018, https://www.technologyreview.com/s/612582/data-that-illuminates-the-ai-boom/. This article relies on Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, Barbara Grosz and Zoe Bauer, "The AI Index 2018 Annual Report", AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA, December 2018, http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf

In the U.S., President Trump noted in his 2019 State of the Union address the importance of investing in cutting-edge industries. The Administration highlighted AI, which relies on advanced data analytics, and announced the launch of "The American AI Initiative."[10] The AI Initiative takes a multipronged approach to accelerating U.S leadership in AI that includes investment, setting governance standards, building an AI workforce, making federal resources available to AI experts, and international engagement. Noting the importance of data and the high expectations for the data-driven market, Canada published its National Data Strategy to address both the economic and non-economic dimensions of advanced data analytics and AI.[11] Singapore announced plans to position itself as Asia's advanced data analytics hub and issued draft guidance for AI oversight.[12] And the European Commission has announced AI ethics guidance that articulates seven requirements that AI systems should meet. This guidance represents part of the European Commission's broader AI strategy to encourage public and private uptake of AI, to ensure member states are prepared for socio-economic changes brought about by AI, and to create an appropriate ethical and legal AI framework.[13]

This grand vision for AI - fueled by advanced data analytics - incorporates plans to realize AI's potential across industry sectors, government and nearly every aspect of society. Advanced data analytics and AI will continue to grow exponentially - encouraged by governments and funded as necessary business strategies.

[10] "Accelerating America's Leadership in Artificial Intelligence," Office of Science and Technology Policy, The White House, February 11, 2019, https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/.
[11] "A National Data Strategy for Canada: Key Elements and Policy Considerations," Centre for International Governance Innovation, CIGI Papers No. 160, February 2018.
[12] "How Singapore plans to become Asia's big data hub in 2018," Singapore Economic Development Board, January 30, 2018, https://www.edb.gov.sg/en/news-and-resources/insights/talent/how-singapore-plans-to-become-asias-big-data-hub-in-2018.html.
[13] "Ethics Guidelines for Trustworthy AI," Independent High-Level Expert Group on Artificial Intelligence (set up by the European Commission), April 8, 2019, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

Growing Recognition of Risks Raised by Advanced Data Analytics

While this vast and growing potential makes advances in the application of analytics inevitable, this progress does not come without risks that potentially compromise individuals' trust. Policymakers, experts and advocates raise concerns that bias may be introduced into algorithms during the knowledge discovery phase and that when the resulting algorithms are applied to individuals in the knowledge application phase discrimination could result. Such discrimination could exclude individuals and groups from participating in activities, receiving benefits or accessing goods and services that should be available to them. Knowledge application also raises issues of personal autonomy and self-determination by suggesting that life decisions individuals have made, for example regarding education and health care, may in the future be made by big data and algorithms. Geo-political developments have highlighted the power of knowledge application and the potential for its abuse when used to segregate groups of individuals in ways that limit their access to legitimate information sources and advance views not based in fact.

Just as policymakers and experts in the U.S., Asia and Europe highlight the economic and social lift promised by advanced data analytics, they also have turned their attention to the risks it may raise.

In the United States, the Congress has taken steps to emphasize and understand the privacy risks presented by advanced data analytics. In February 2019, the House Committee on Energy and Commerce held a hearing titled "Protecting Consumer Privacy in the Era of Big Data."[14] In his prepared opening statement, Chairman Frank Pallone noted the vast array of benefits advanced data analytics can yield. But he also commented on the data collected about individuals through technology and the powerful inferences about them that advanced data analytics can reveal. He voiced concerns about discrimination, differential pricing and physical

[14] https://energycommerce.house.gov/committee-activity/hearings/hearing-on-protecting-consumer-privacy-in-the-era-of-big-data.

harm that could result from uses of data, noting that data is collected and processed in an environment where few rules apply.[15]

Also, in the United States, advocates from over 40 civil rights, civil liberties, and consumer groups have called on Congress to address discrimination that can result from the Internet economy. In a letter to Congress,[16] they called on legislators to protect civil rights, equity, and equal opportunity in the digital environment. The organizations wrote that any privacy legislation must be consistent with the "Civil Rights Principles for the Era of Big Data," issued in 2014.[17]

Policymakers outside the U.S. also have voiced concerns about risks arising from advanced data analytics and the need to address them if benefits are to be realized. Canada's national data strategy cites data's "enduring and uniquely potent influence on individual lives, social relationships and autonomy." It references data's implications not only for commerce but for the operation of democracy. Any data strategy, it cautions, would need to address both the economic and non-economic dimensions of advanced data analytics, balancing goals that included reaping economic gains, respecting privacy, preserving an open society and democracy, maintaining public security and building institutions that maintain or enhance Canada's national identity.[18] Similarly, the EU's "Ethics Guidelines for Trustworthy AI" acknowledge that while AI offers substantial benefits to individuals and society, it also poses certain risks and may have negative consequences, including some which may be difficult to anticipate, identify or measure (e.g. on democracy, the rule of law and distributive justice or on

[15] "Pallone Remarks at Data Privacy Hearing," House Energy and Commerce Committee Newsroom, Chairman Frank Pallone, Jr., 116th Congress, February 26, 2019. https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/0226%20FP%20Opening%20Statement%20CPC%20Data%20Hearing.pdf.

[16] http://civilrightsdocs.info/pdf/policy/letters/2019/Roundtable-Letter-on-CRBig-Data-Privacy.pdf

[17] These include: stopping high-tech profiling, ensuring fairness in automated decisions, preserving constitutional principles, enhancing individual control of personal information, and protecting people from inaccurate data. "Civil Rights Principles for the Era of Big Data," The Leadership Conference on Civil and Human Rights, February 2014, https://civilrights.org/2014/02/27/civil-rights-principles-era-big-data/.

[18] High-Level Expert Group on AI, "Ethics Guidelines for Trusted AI," April 2019, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

the human mind itself).[19] In response, the Guidelines provide that when appropriate, organizations must adopt mitigation strategies proportional to the risks.[20]

If the benefits of advanced data analytics are to be realized, it is critical that these concerns about risk are addressed. Guidance in law and policy lays the foundation for the public's trust in data analytic activities. It also provides organizations with the certainty needed to allow them to use advanced data analytics with confidence to derive and apply the insights data can yield. But to be optimally effective, guidance must be informed by the realities of how advanced data analytics work in practice.

Data Protection Legislative and Policy Response

As advances in advanced data analytics continue, policymakers and governments are taking steps to implement data protection laws and guidance. In the European Union, for example, the GDPR includes provisions governing automated decision-making, including profiling, prohibiting such data processing except when certain conditions apply. More recently, the European Commission issued guidance for organizations implementing AI. The guidance requires that organizations use AI for ethical purposes, in accordance with fundamental rights, principles and values. It derives guidance about the realization of trustworthy AI with respect both to ethical purpose and technical robustness. Finally, it articulates a concrete assessment list based on these requirements to assist organizations.[21]

Moreover, policymakers, advocates, businesses and law makers in the United Sates have renewed calls for privacy legislation. The Trump administration is developing consumer data privacy policies; the Commerce Department has met with technology companies as it anticipates how these policies will be enshrined in legislation.[22] The Commerce Committee of

[19] Ibid.
[20] Op. Cit., fn. 12.
[21] Ibid.
[22] Shepardson, David, "Trump administration working on consumer data privacy policy," *Reuters,* July 27, 2019, https://www.reuters.com/article/us-usa-internet-privacy/trump-administration-working-on-consumer-data-privacy-policy-idUSKBN1KH2MK.

the U.S. Senate recently held a hearing titled "Policy Principles for a Federal Data Privacy Framework in the United States,"[23] and the Federal Trade Commission recently held a hearing on the "FTC's Approach to Consumer Privacy."[24] Legislators have proposed bills that would address privacy generally; consumer and privacy advocates have also released draft legislation. U.S. companies and trade associations have issued their own proposals. Data privacy solutions have taken the form of high-level frameworks, guidance documents and draft legislation. In some cases, these proposals anticipate the use of data for analytics; others reflect older notions of data collection and processing.

However, legislation – whether proposed, enacted or implemented - has not resolved the conundrum inherent in advanced data analytics governance. It has not addressed the repurposing of data that is essential to advanced data analytics. The GDPR, for example, establishes research as a compatible purpose but does not define what constitutes research for purposes of the GDPR. And while the GDPR suggests a risk-based approach to governance, its analysis places far greater emphasis on risk than benefits. Thus, the concept of legitimate interests established in the GDPR would seem to be the means for lawfully conducting data driven knowledge creation, but guidance has yet to make that clear or explain how it might be applied in cases of advanced data analytics.

Data protection and privacy law and guidance should provide the foundations for public trust and legal certainty needed to allow advanced data analytics to reach its full potential. But to do so, legislative solutions must be grounded in an understanding of how advanced data analytics, including big data, works, the benefits advance data analytics promise, where the risks do and do not arise, and how to address those risks. Failure to design thoughtful, workable solutions in law could unduly limit the possibilities of advanced data analytics.

---

[23]Hearing of the U.S. Senate Commerce Committee, February 27, 2019, https://www.commerce.senate.gov/public/index.cfm/2019/2/policy-principles-for-a-federal-data-privacy-framework-in-the-united-states

[24] Hearing of the U.S. Federal Trade Commission, April 9, 2019, https://www.ftc.gov/news-events/events-calendar/ftc-hearing-competition-consumer-protection-21st-century-february-2019.

Revisiting the Knowledge Discovery/Knowledge Application Distinction

In order to create sound governance for the risks raised during advanced data analytics, it is necessary to understand the knowledge discovery/knowledge application distinction. Perhaps the most important contribution of the 2013 Paper is its characterization of advanced data analytics as a two-phase process: *knowledge discovery* and *knowledge application*. Distinguishing between knowledge discovery and knowledge application makes it possible to determine at what point in processing risk occurs, the nature of the risk, and how it can be mitigated. Describing advanced data analysis in this way also provides the foundation for governance that allows for robust exploration of data to determine what insights can be gleaned and what protections for individuals are needed when insights are applied. Therefore, the nature of this distinction deserves review.

Knowledge Discovery

In the knowledge discovery phase, data scientists 1) format data sets, 2) analyse them to discover what the data may reveal, and 3) interpret the analysis to understand how conclusions were reached and whether they are scientifically sound.[25] For ease of understanding, these phases are characterized as discrete and separate. In practice, however, the phases of knowledge discovery often overlap.

The product of knowledge discovery is an algorithm, which can perform a variety of tasks: classifying discrete variables; calculating continuous variables (such as the value of a home based on its attributes and location); segmenting data into groups or clusters; identifying correlations between different attributes in a data set; and sequence analysis.

It is important to note the differences between the scientific method and knowledge discovery. The scientific method consists of identifying a question, developing hypotheses and conducting

[25] The 2013 paper outlines the steps involved in knowledge discovery more specifically. These include: data *acquisition*, *preprocessing* data into a consistent format that can be analyzed; *integration* and consolidation of data for analysis; *analysis*, or searching for relationships, classifications or associations between data items in a database; and *interpretation*. Op cit., fn. 1 at pp. 9-10.

tests to test those hypotheses. The scientific method also involves testing theories of causation – identifying a variable that influences other variables or causes other things to occur. In contrast, knowledge discovery is not conducted with a specific hypothesis to be tested.[26] Rather, it involves sifting, reviewing and analyzing data to determine what predictions, correlations or inferences may be revealed. What emerges from knowledge discovery may be unexpected or appear, at least initially, unrelated to the kinds of data analyzed. Further, knowledge discovery reveals correlations rather than causes. Correlations are useful because they can indicate a predictive relationship that can be exploited in practice.[27]

Knowledge Application

In the knowledge application phase of advanced data analytics, the algorithm derived in the knowledge discovery phase is used to make decisions based on predictions or inferences that result from knowledge discovery. In other words, based on what they learn in knowledge discovery, organizations act to make decisions that affect individuals and groups of individuals. This, in turn, may have societal consequences as well. For this reason, the knowledge application phase is viewed as the aspect of advanced data analytics that hold the greatest risk.

This distinction between knowledge discovery and knowledge application holds implications for advanced data analytics governance. The model proposed in the 2013 Paper envisioned a break in data analysis between the two phases that would provide an opportunity to understand the knowledge created (i.e., the algorithm), assess any risk arising from its application, and mitigate that risk. The breakdown in this model is particularly evident when big data analytics are used in AI, where the pace of processing is rapid – often occurring in real time – and there is no natural break between knowledge discovery and knowledge application.

[26] It is for this reason that organizations engaging in knowledge discovery cannot rely on consent as a legal basis to process data. Because the insights data hold are not revealed until the data are analyzed, consent to processing cannot be obtained based on an accurately described purpose.

[27] For example, an electrical utility may produce less power on a mild day based on the correlation between electricity demand and weather. In this example, there is a causal relationship, because extreme weather causes people to use more electricity for heating or cooling. However, in general, the presence of a correlation is not sufficient to infer the presence of a causal relationship (i.e., correlation does not imply causation).

New data is introduced to algorithms which may, in turn, generate new algorithms without the benefit of risk assessment and mitigation. Governance for advanced data analytics must take into account this rapid pace and seamless generation of algorithms in AI.

Advanced Data Analytics and Governance

The 2013 Paper considered at some length the challenges advanced analytics raise for traditional approaches to data protection. In particular, some longstanding notions of fair information practices raise significant governance and compliance hurdles for organizations applying advanced data analytics. For example, while individual consent has long been central to data protection, the ubiquitous and ongoing collection of data often makes consent unwieldy or impossible. Because advanced data analytics may serve purposes that only become evident through knowledge discovery, organizations at the time of collection either may not be able to describe to what purpose data will be put or will be forced to articulate purposes so broadly that the notices will lose their meaning and usefulness. Further, the need for vast, diverse data stores for knowledge discovery argues against the principle of data minimization, which requires that organizations collect only the data necessary and relevant to carrying out a specified purpose and that data be disposed of when it is no longer needed for that purpose. And as noted previously, the repurposing of data – using it for a purpose not anticipated when initially collected – challenges the principle of purpose specification.[28]

Such challenges led the IAF to ground its work on advanced data analytics governance on the fair information practice principle of accountability, applying the essential elements to decision-making about advanced data analytics. By distinguishing between the knowledge discovery and the knowledge application phases of analytics, IAF guidance helps organizations identify where risks arise and determine whether and to what extent they can be mitigated. By applying ethics principles, the guidance takes into account not only the full range of risk, but equally important,

[28] The OECD principle of purpose specification states: "The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose."

the full range of benefits to individuals that may result from processing. In doing so, IAF guidance allows organizations greater latitude to use data for beneficial purposes but holds them to account for making responsible decisions about data use and for refraining from using data when the risk to individuals is too great or cannot be mitigated adequately.[29]

Legitimate Interests – A Basis for Guidance for Advanced Data Analytics

The IAF believes that the concept of legitimate interests can form the basis for governance of the *knowledge discovery* of advanced data analytics. Originally articulated in European law, the concept is now reflected in the data protection regime in Brazil, Colombia, and will most likely e reflected in updated laws one might see in Canada, Argentina, and other countries as well. The legitimate interest basis for processing often is relied upon when data collected for one purpose is used for another, not inconsistent purpose.[30] Such repurposing is essential to advanced data analytics, where large troves of data, gathered from disparate sources and for a range of uses, are processed in knowledge discovery.

Applying the Legitimate Interests Test

Legitimate interest is articulated in the GDPR as one of six legal bases for processing data. An organization that wishes to rely on legitimate interest must conduct an assessment that balances its legitimate interest in using data against the risk that use may raise to the individual's fundamental rights and freedoms. The GDPR also requires that data protection be conducted in a risk-based manner, whereby the organization determining whether data processing falls within its legitimate interest weighs not only the risks but also the benefits to stakeholders.[31]

---

[29] "Legitimate Interests and Integrated Risks and Benefit Assessment: A Framework for Determining if Processing as Permitted by Legitimate Interests is Legal, Fair and Just," The Information Accountability Foundation, http://informationaccountability.org/wp-content/uploads/Legitimate-Interests-and-Integrated-Risk-and-Benefits-Assessment.pdf.

[30] The Article 29 Working Party's "Opinion on the notion of legitimate interests of the data controller under Article 7 of the Directive 95/46/EC" is a plentiful source of examples of repurposing of data. https://www.huntonprivacyblog.com/wp-content/uploads/sites/28/2014/04/wp217_en.pdf

[31] The IAF and the UK Data Protection Network in its "Legitimate Interests Guidance" have suggested assessment processes that take into account both benefits as well as risks. https://www.dpnetwork.org.uk/dpn-legitimate-interests-guidance/.

The IAF believes that, as a general matter, use of advanced data analytics *for knowledge discovery* would fall within the contours of the legitimate interest basis for processing, if appropriate safeguards are in place. Because of the potential value derived in knowledge discovery and the arguably low risk advanced data analytics raise for individuals when used for knowledge discovery, legitimate interest could supply the necessary legal basis for processing. However, organizations still must conduct the requisite analysis to determine whether any particular instance of knowledge discovery meets the requirements of legitimate interests. The IAF believes that specific guidance, based on the principle of accountability and related guidance, is still needed to help organizations conduct that analysis.

Challenges to a Legitimate Interests Approach

The Credibility of the Organization as Decision-maker

The concept of legitimate interests is, however, subject to criticism: how can the determination that the use of data for an organization's legitimate interests be trusted, when the organization itself conducts the analysis? Reliance on legitimate interests raises appropriate scepticism that decisions related to individuals conducted by organizations (private, academic and public) will be arrived at in an honest, fair and competent manner. Addressing that scepticism, when data is used for knowledge discovery, requires 1.) guidance about how legitimate interest analysis should be carried out and the appropriate criteria to be evaluated and 2.) the nature of oversight necessary for the assessment process. The IAF, building on its own work and its work with the Indiana University Center for Law, Ethics, and Applied Research, will explore how data ethics research boards, whether located within or outside of organizations, can provide oversight to assessments for whether knowledge creation is conducted in an appropriate manner. To establish necessary credibility, the IAF recognizes the need to develop guidance - based in accountability - that promotes the competency, integrity and independence of the research.

The Individual's Right to Object to Processing

Under Article 6(1) of the GDPR, individuals have the right to object to processing undertaken on the basis of an organization's legitimate interests. The IAF believes that while such a right is appropriate in the knowledge application phase of advanced analytic processing - when decisions are made about individuals – exercise of such a right may compromise the accuracy of insights, and therefore lead to incorrect to decisions that may negatively affect other stakeholders. While the right to object to processing may be settled law in the EU, policymakers should consider carefully whether the right to object should apply in knowledge discovery, where the risk to individuals is low. This determination will rely, at least in part, on the requirement that organization's credibly balance of stakeholder interests, and assess and mitigate risks that may arise from the advanced data analytics processing required for knowledge discovery.

Restrictions on Use of Sensitive Data

Article 9 of the GDPR prohibits, with certain exceptions, processing of sensitive data, defined in the GDPR to include, among others, racial or ethnic origin, processing of genetic data, biometric data and data concerning health or sexual orientation. While the IAF recognizes the concerns raised by processing of such data and the need for appropriate safeguards, it also understands that these prohibitions could negatively affect or impede knowledge discovery related to health research. While Article 9 provides an exception to this prohibition when individuals' consent has been obtained, that consent may be withdrawn. To arrive at the most accurate, representative and unbiased results in knowledge discovery, data scientists rely on the most robust, comprehensive data sets. Withdrawal of data could affect the results of knowledge discovery, leading to flawed or discriminatory outcomes when those results are applied.

Conclusion

The work undertaken by the IAF to date suggests that questions about when the use of advanced data analytics for knowledge discovery is trustworthy and appropriate can be

supported by a legitimate interest analysis. In addition, as a result of this work, IAF believes that because of the broad benefits promised by advanced data analytics, weighing three interests – benefits to the individual and society, the legitimate interest of the organization, and risks to the right of individuals that may arise from processing - will be critical to accountability-based governance of advanced data analytics.

For these reasons, the IAF recommends that privacy and data protection law, regulation and guidance that governs advanced data analytics should:

- be based on its two-phased process – *knowledge discovery* and *knowledge application;*
- recognize the nature and level of risk each phase raises, and the risks inherent when the decision is made not to process data for knowledge;
- introduce "legitimate interests" as a basis for lawful processing in knowledge discovery;
- accountability-based guidance for how a legitimate interest analysis should be carried out that provides clarity and certainty for organizations, particularly when special categories of data are used for knowledge discovery; and
- take into account the rapid pace and seamless generation of algorithms in AI.

IAF looks forward to continuing to work with stakeholders to develop a robust, credible and workable approach whereby organizations can establish legitimate interests as a legal basis for advanced data analytics knowledge discovery.