Enforcing Big Data Assessment Processes

IAF Big Data Ethics Initiative, Part C

Executive Summary

The Information Accountability Foundation ("IAF") released "A Unified Ethical Frame for Big Data Analysis" ("Unified Ethical Frame" or "Frame") at the International Conference of Data Protection and Privacy Commissioners in October 2014.[1] The Unified Ethical Frame defines an assessment process to determine whether big data undertakings are legal, fair and just and to demonstrate how that determination was reached. The Frame is the first of four projects in the IAF's big data ethics initiative. The subsequent two projects create a common and a customised assessment framework. This paper is the product of the final project, a consideration of how these assessment frameworks might be overseen by data protection authorities ("DPAs")[2] in various jurisdictions including Europe, the Americas, Asia Pacific and Africa ("Enforcement Paper Project"). This paper is intended to present issues and encourage discussion about how such oversight might be designed, even in the absence of clear legal guidance specific to big data analytics.

Opinion 4/2015, "Towards a New Digital Ethics," issued by the European Data Protection Supervisor on 11 September 2015 ("the EDPS Opinion")[3] reflects the need for such ethical assessments. The EDPS Opinion calls for controllers to be fully accountable, particularly when using big data. Accountability requires an assessment of the risk to individuals and documentation of how that risk is assessed and mitigated. The Unified Ethical Frame describes a process for reaching a determination based on recognised criteria about whether a particular big data analytics project is ethical and appropriate.

The IAF worked with various DPAs in developing this paper. The IAF especially thanks the Agencia Española de Protección de Datos ("Spanish AEPD" or "AEPD") and the Garante Per La Protezione Dei Dati Personali ("Italian Garante" or "Garante") for hosting workshops in Madrid and Rome. IAF staff learned a great deal in working with the various DPAs that participated in the Enforcement Paper Project.

As initially conceived by the IAF, the internal big data assessment process and the governance around that process would be a code of conduct for big data assessment, and DPAs would have jurisdiction to enforce against organisations that did not fulfil their obligations pursuant to the code. As Enforcement Paper Project participants explored that concept, it became apparent that governing laws and traditions around the world interpret codes of conduct in different ways.[4] Any solution, therefore, would have to embrace the diverse approaches taken by

[1] The Information Accountability Foundation (2014) "A Unified Ethical Frame for Big Data Analysis", http://informationaccountability.org/wp-content/uploads/IAF-Unified-Ethical-Frame.pdf.

[2] While data protection and privacy protection agencies and authorities have many different structures and names, for the purpose of this paper, they will be collectively referred to as DPAs.

[3] EDPS (2015), "Opinion 4/2015: Towards a New Digital Ethics", https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-09-11_Data_Ethics_EN.pdf.

[4] The European Union is currently considering a draft data protection regulation that contains an article 38 that addresses the means for creating and supervising codes of conduct.

various legal systems and work toward interoperability across those systems and cultures.

Further, the big data assessment process ideally should be considered one part of an organisation's overall information accountability governance programme, specifically customised to address big data analytics. Organisations using the assessment process should be transparent about their use of an assessment process and be prepared to demonstrate the efficacy of their big data governance programmes and processes to supervisory bodies, e.g., DPAs or accountability agents overseen by DPAs.

Legal requirements for oversight of big data assessment processes differ from jurisdiction to jurisdiction. A fully acknowledged assessment process might be directly enforced in some jurisdictions as a code of conduct, and in other jurisdictions, it might serve as a proxy for compliance with law or regulations. In jurisdictions where regulators are ombudsman, excellent practices for big data assessment could be suggested as an addendum to accountability guidance, and incidents could be reviewed based on the presence of such an assessment process.

The discussions highlighted that many issues related to big data assessment would benefit from clear consensus on definitions. For example, when is research considered scientific research, and when is scientific research considered a compatible data use?

In the end, Enforcement Paper Project participants found the following:

- Local legal mandates and DPA attributes will define the nature of the oversight mechanisms. Some jurisdictions may recognise assessment processes as a code of conduct, while other jurisdictions may consider them to be a best practice.
- When assessing big data analytics, it is important to differentiate the discovery phase of the analytics from the implementation phase. The discovery phase involves analysing diverse data sets to arrive at correlations and insights. The discovery phase often reveals powerful insights that may be sensitive and that individuals do not expect. Moreover, the discovery phase, may involve uses of data for purposes that were not envisioned when the data was originally collected. However, the discovery phase typically does not have direct impact on individuals. The application phase involves using insights to make decisions. The application phase is where impact on individuals is more likely to occur and therefore is more easily governed by traditional data protection methods. Risks and benefits of analysing data should be assessed as part of both phases.
- A legal basis, such as legitimate interests, is necessary to govern big data and big data assessment processes. Data protection and privacy legal systems that have legitimate interests as a legal basis for processing have a governance advantage. Where legitimate interest, or something similar to it, does not exist, the legal permission for big data processing, especially discovery, will be more difficult. Furthermore, the balancing process developed for big data has application for the broader field of legitimate interests.

- It will be important to reach a broad consensus on when data-driven research is considered scientific research. The definition of scientific research will be particularly important in determining when data may be used for research purposes during the discovery phase of big data analytics.
- This paper does not fully discuss the issues related to certification agencies. However, resolving the issues related to such accountability agents may be material to using codes of conduct in the oversight of uses of big data.
- Standing ready to demonstrate is neither a purely self-regulatory nor a passive process. In general, accountable organisations using this process should have a methodology for charting governance and assessments pursuant to a code of conduct so that a full story might be provided to the oversight agent at any point of time with minimal notice. Companies such as Nymity have provided services for charting accountability and identifying the evidence necessary to demonstrate accountability. Similar services will be necessary for big data codes as well.

The IAF issues this paper for the sole purpose of sparking a discussion amongst the DPA community. Any shortcomings found in this paper should be attributed to IAF staff, not to participants in or funders of the Enforcement Paper Project.

Introduction

In late 2013, the IAF convened industry representatives, experts and former DPAs[5] to develop a process and criteria data stewards could use to determine whether a big data discovery or application process is appropriate or not. This big data initiative has produced three deliverables: the Unified Ethical Frame published in October 2014,[6] the Big Data Assessment Process issued in March 2015,[7] and a specific assessment process customised for the digital marketing industry that began testing in May 2015[8].

The Unified Ethical Frame lays out the basic philosophy for big data governance. The foundation of the Frame is a governance process that balances the interests of the individual, society and the organisation conducting a big data project. This balancing process considers the full range of fundamental rights and interests that impact the individual, including the fundamental rights of human dignity, privacy, and data protection. Supplemental to the Frame is a common assessment framework that considers and balances the various interests. While this framework can be customised to meet the needs of organisations, industries and sectors, the basic assessment framework provides the key question areas that must be addressed.

At the very outset, the IAF research team determined that big data assessments could not stand alone. Rather, they should be incorporated into an organisation's information accountability governance programmes and processes that comply with the legal requirements applicable to the organisation. Such big data assessments complement these existing accountability programmess by creating a framework for assessing the risks and benefits for all stakeholders raised by big data analytical processing and by isolating risks that must be mitigated.[9] The process emphasises the risks and benefits to the individuals to whom the data pertains.

Determinations that big data analytics are legal, fair and just will only be trusted and fully effective if the assessment process that demonstrates how these determinations were reached have enforcement. Thus, big data governance processes should be subject to external oversight. Such oversight could require organisations to stand ready to demonstrate the programme on request to DPAs or accountability agents overseen by DPAs.

The IAF's initial discussions with DPAs focused on the procedural steps necessary to make codes of conduct fully operational. Those steps included the following:

---

[5] The members of the research team are listed on Appendix A to the Unified Ethical Frame.
[6] The Information Accountability Foundation (2014).
[7] The Information Accountability Foundation (2015) "Big Data Assessment Framework", http://informationaccountability.org/wp-content/uploads/IAF-Big-Data-Ethics-Initiative-Part-B.pdf
[8] The Information Accountability Foundation (2015) "Contextual Assessment Worksheet for the Marketing Industry", http://informationaccountability.org/wp-content/uploads/IAF-Big-Data-Ethics-Initiative-Part-D-For-Marketing.pdf.
[9] Centre for Information Policy Leadership (CIPL) (2009), "Data Protection Accountability: The Essential Elements", http://www.huntonfiles.com/files/webupload/cipl_galway_accountability_paper.pdf.

- Linking codes of conduct to a DPA's legal mandate, i.e., does the DPA have the legal authority to enforce a code of conduct
- Establishing a code of conduct or adopting a code of conduct by a company.
- Publishing a description of how individuals and organisations benefit when codes of conduct.
- Developing a mechanism by which a DPA receives a company's attestation that it will abide by a code of conduct
- Establishing a means for individuals to complain if they believe the code of conduct is not performing as promised
- Developing a mechanism for DPAs to randomly check an organisation's compliance with a code of conduct
- Creating a mechanism for companies to renew an attestation

The steps above are similar to those required for Binding Corporate Rules (BCRs) in the European Union (EU) and Cross-Border Privacy Rules (CBPRs) in the Asia-Pacific region for governing data transfers. However, those procedural steps had been established based on a previously agreed upon policy consensus.[10]

In order to discuss the operationalisation of codes of conduct, the DPAs that IAF consulted with wanted a clear understanding of what would be overseen and how it would fit into a comprehensive privacy programme. In response, IAF articulated, the elements of a big data code of conduct that would be subject to DPA oversight ("Box 1" below). Fundamentally, a code of conduct is an organisation's internal policy mandate that involves an assessment process. It comprises an organisation's mechanism for demonstrating its process to assess the full set of risks associated with big data processing, its strategy to mitigate those risks, the basis for arriving at those decisions, its review to determine whether its risk mitigation is effective and on going internal oversight. The assessment framework, a data protection impact assessment customised for big data, is the mechanism for the demonstration.

Box 1. Elements of a Big Data Code of Conduct

1. Internal policy that mandates assessments and their integration in internal governance.
2. Assessment tool that contains the elements of the assessment framework.
3. Decisions, mitigations and evidence used to make those conclusions.
4. Internal oversight over the big data process.
5. Standing ready to demonstrate the process.

[10] It was quickly determined that describing procedural steps for big data codes of conduct was premature and that basic policy questions needed resolution first.

A code of conduct can be either an internal instrument that mandates internal ethics but is subject to external review or an external industry instrument.[11]

External Oversight: Five Questions for Consideration

Discussions with the DPAs revealed that identifying an appropriate external oversight mechanism requires consideration of the following questions:

1. What is being overseen?
2. Who will oversee it?
3. Under what authority?
4. For what purpose?
5. With what outcome?

These five questions are reviewed in turn.

1. What is Being Overseen?

The first project, the Unified Ethical Frame, specifies that the assessment process must strive to identify how the various interests of stakeholders, particularly individuals, have been adequately balanced. To achieve that end, the Unified Ethical Frame sets forth five core values against which to assess. Those five values act as a compass for the common and customised assessment frameworks. The five values are as follows: beneficial, progressive, sustainable, respectful and fairness.

The assessment framework is based on these five values and is designed to raise the issues necessary to make sound ethical judgments as they concern new big data analytics. The assessment framework is organised into sections with the five values reflected throughout the sections. Since familiarity with the five key values is essential, this paper repeats them almost in their entirety from the Unified Ethical Frame.

1a. Five Core Values

Beneficial

Both the discovery and application phases require an organisation to define the benefits that will be created by the analytics and should identify the parties that gain tangible value from the effort. The act of big data analytics may create risks for some individuals and benefits for others or society as a whole. Those risks must be counter-balanced by the benefits created for individuals, organisations, political entities and society as a whole. Some might argue that the creation of new knowledge is a value-creating process itself. While big data does not always begin with a hypothesis, it

[11] For example, the Federation of European Direct Marketing Associations has had an approved data protection code of conduct for years.

usually begins with a sense of purpose about the type of problem to be solved. Data scientists, along with others in an organisation, should be able to define the usefulness or merit that comes from solving the problem so that it might be evaluated appropriately. The risks should also be clearly defined so that they may be evaluated as well. If the benefits that will be created are limited, uncertain, or if the parties that benefit are not the ones at risk from the processing, those circumstances should be taken into consideration and appropriate mitigation for the risk should be developed before the analysis begins.

Progressive

Because bringing large and diverse data sets together and looking for hidden insights or correlations may create some risks for individuals, the value from big data analytics should be materially better than not using big data analytics. If the anticipated improvements can be achieved in a less data-intensive manner, that less intensive processing should be pursued. One might not know the level of improvement in the discovery phase. Yet, in the application phase, the organisation should be better equipped to measure it. This application of new learning to create materially better results is often referred to as innovation. There are examples of big data being used to reduce congestion, manage disaster relief and improve medical outcomes. These are all examples of material improvements; however, there are other examples where organisations may analyse data and achieve only marginal improvements but use big data, because big data is new and interesting. Organisations should not create the risks associated with big data analytics if there are other processes that will accomplish the same objectives with fewer risks.[12]

Sustainable

All algorithms have an effective half-life–a period in which they effectively predict future behaviour. Some are very long; others are relatively short. The half-life of an insight affects sustainability.

Big data analysts should understand this concept and articulate their best understanding of how long an insight might endure once it is reflected in application. Big data insights, when placed into production, should provide value that is sustainable over a reasonable period. Considerations that affect the longevity of big data analytics include whether the source data will be available for a period of time in the future, whether the data can be kept current, whether one has the legal permissions to process the data for the particular application and whether the discovery may need to be changed or refined to keep up with evolving trends and individual expectations.

There are situations where data, particularly de-identified data, might be available for

[12] Data protection guidance often raises the issue of proportionality. Those concepts of proportionality come into play when conducting assessments on all the values, but they particularly come into play on progressive.

the discovery phase but would not be available in the application phase because of legal or contractual restrictions. These restrictions affect sustainability.

Respectful

Respectful relates directly to the context in which the data originated and to the contractual or notice related restrictions on how the data might be applied.

- The United States Consumer Privacy Bill of Rights speaks to data being used within context.
- European law discusses processing not incompatible to its defined purpose.

Big data analytics may affect many parties in many different ways. Those parties include individuals to whom the data pertains, organisations that originate the data, organisations that aggregate the data and those that might regulate the data. All of these parties have interests that must be taken into consideration and respected. Organisations using big data analytics should understand and respect the interests of all the stakeholders involved in, or affected by, the application. Anything less would be disrespectful.

Fairness

Fairness relates to the insights and applications that are a product of big data, while respectful speaks to the conditions related to, and the processing of, the data. In lending and employment, United States law prohibits discrimination based on gender, race, genetics or age. Yet, big data processes can predict all of those characteristics without actually looking for fields labelled gender, race or age. The same can be said about genotypes, particularly those related to physical characteristics. Section 5 of the United States Federal Trade Commission ("FTC") Act prohibits unfair practices in commerce that are harmful to individuals not outweighed by countervailing benefits.[13] European guidance on application of the data protection directive continually references fairness as a component of determining whether a use of data is incompatible or a legal basis to process is appropriate. Big data analytics, while meeting the needs of the organisation that is conducting or sponsoring the processing, must be fair to the individuals to whom the data pertains.

The analysis of fairness needs to look not only at protecting against unseemly or risky actions but also at enhancing beneficial opportunities. Human rights speak to shared

[13] FTC Policy Statement on 17 December 1980 states: (1) whether the practice, without necessarily having been previously considered unlawful, offends public policy as it has been established by statutes, the common law, or otherwise-whether, in other words, it is within at least the penumbra of some common law, statutory or other established concept of unfairness; (2) whether it is immoral, unethical, oppressive or unscrupulous; (3) whether it causes substantial injury to consumers (or competitors or other businessmen). U.S. Federal Trade Commission (1980), "FTC Policy Statement on Unfairness", http://www.ftc.gov/public-statements/1980/12/ftc-policy-statement-unfairness.

benefits of technology and broader opportunities related to employment, health and safety. Interfering with such opportunities is also a fairness issue.

In conducting this fairness assessment, organisations should take steps to balance individual interests with integrity.

Implicit in all five values is the openness required the by fair information practice principles.[14] Organisations conducting big data analytics must find ways to communicate what they do and how they do it. Furthermore, they should look for ways for openness to be effective. In some instances, this might be within privacy notices. In other cases, it might be just-in-time notices, dashboards, videos or other means that are effective. Organisations should stand ready to demonstrate means for making visible to stakeholders how big data analytics are designed and operated.

1b. When the Assessment Framework Should be Used

Organisations should implement the assessment framework as big data projects reach key milestones or decision points. When these events will occur will vary based on the nature of the industry, the sector, and the organisation. The assessment framework should be integrated into the established practices of organisations.

The preamble to the assessment framework provides for four implementation points across the big data analytics life cycle: in addition to the discovery and application phases in the middle, there is the concept phase at the beginning and the review phase at the end of the big data analytics life cycle.

Concept

Organisations articulate their reasons for using specific data sets, document the new data created by the analytics, identify the opportunities the analytics offer for new insights and describe the potential utility of those insights and the possibilities for their further application before any real analytics takes place. The results of this reasoning are presented to decision makers who will then determine whether to proceed to the discovery phase.

Discovery

During the discovery phase, data scientists analyse data sets to understand what new insights the data might reveal. The discovery phase involves the acquisition, formatting and consolidation of data prior to its analysis and interpretation. While the assessment process will vary based on sector and industry, at some organisations, this pre-analysis will take place at the concept phase.

[14] U.S. Federal Trade Commission (1998), "Privacy Online: A Report to Congress", https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf.

The questions raised during the concept phase are also relevant to the discovery phase.[15] However, it is unnecessary to ask these questions during both phases. Rather, they are posed at a point during the assessment process appropriate to the nature of the big data analytics so that in some cases the questions will be asked during the concept phase and in other cases they will be asked during the discovery phase. In some instances, the questions may need to be duplicated because of specific sensitivities.

Application

A decision to move forward or not is made between the completion of the discovery phase and the beginning of the application phase. The organisation must think beyond its interests to the interests of the other stakeholders. The organisation must determine whether the analytics will create real benefits and who will receive those benefits, apart from the objectives or interests of the organisation. The organisation further must determine whether the insights will be sustainable once analytics commence; whether improvements in analytics are significant enough to justify more robust big data analytics; and whether the application is respectful and fair. Much of this evaluation may have taken place at the concept and discovery phases, and if that reasoning is still relevant, it need not be repeated. Consideration is given as to whether key questions have been asked and answered. The decision maker is responsible for the integrity of the process.

Review

In order to determine whether internal programmes and processes that govern decision making about the use of big data are effective, organisations should conduct on going reviews. An ethical review takes place when customary inspections of new applications of data are scheduled. The level of the ethical review is proportional to the constant evolution of the programmes. New data sets may have been introduced, or processing shortcuts may have been developed. If changes are extensive, the ethical review is similarly robust. Warning signals, such as complaints by individuals related to outcomes, are taken seriously. The Unified Ethical Frame is not about generating additional work but creating internal controls, where necessary, and doing so with integrity.

1c. The Purpose of the Assessment Framework

The purpose of the assessment framework is to aid an organisation's judgment about the appropriateness of big data analytics. A customised assessment framework is intended to

[15] For example, what are all the sources of the data? Can the source data be kept current over time? If not, is there an adequate replacement? Are there legal, policy, contractual, industry, or other obligations linked to the data? Is the data linkable to a particular individual or not? Who are all the stakeholders related to both the analysis and the use of the resulting insights?

identify issues for consideration during an ethical review. The assessment framework does not generate automatic decisions for users. Instead, the assessment framework helps decision makers incorporate what they learn from the review process into their decision making. Documentation of the review process provides evidence that supports the decisions made.

The assessment framework supports an evaluation of each phase of analytic processing to determine whether the analytics are beneficial, sustainable, progressive, respectful and fair. In turn, the assessment framework promotes judgment, based on the risks, interests, rights and benefits of all the various inside and outside stakeholders. The assessment framework assists organisations in arriving at the right balance as they make decisions about big data analytic processing.

Big data analytics do not fit into a black-or-white governance model. The goal of the assessment framework is to assist an organisation in determining whether its big data analytics are legal, fair and just and to demonstrate how that determination was reached. The assessment framework helps the user achieve responsible decision making.

2. Who Will Oversee?

Existing governance and oversight processes in the EU and the Asia Pacific region could serve as models for successful external oversight of big data analytics. While these two regions share a similar objective – trustworthy governance for cross-border data flows – each of them employs different devices to achieve that objective.

The Article 29 Data Protection Working Party established under Directive 95/46/EC ("Article 29 Working Party") provides for the oversight of BCRs, codes of conduct for the international transfer of personal data within the same corporate group subject to the authorisation of the relevant DPAs.[16] BCRs adopted by a corporate group must create duties for the group to have data protection audits and to cooperate and comply with DPAs regarding the BCRs and must allow DPAs access to the audit results and give them the authorisation and power to carry out a data protection audit themselves. Thus, BCRs are directly overseen by the DPAs.

In the Asia Pacific region, the Asia-Pacific Economic Cooperation ("APEC") economies endorsed the APEC Privacy Framework which established the CBPRs. CBPRs govern the movement of data across the Asia Pacific region. Accountability agents identified by governments and approved by an APEC-wide governing process oversee the efficacy of the CBPRs and their implementation. DPAs provide back-up enforcement to the oversight of CBPRs by accountability agents.

[16]The Article 29 Data Protection Working Party is composed of all the Data Protection Authorities in the Member States, the EDPS and the European Commission. Article 29 Data Protection Working Party (2008), " Working Document Setting up a framework for the structure of Binding Corporate Rules ", WP 154, http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2008/wp154_en.pdf.

Big data oversight could work similar to BCRs and CBPRs. To be effective, big data oversight must take into account the legal and cultural systems across the different regions or even sub-regions where it applies. Codes of conduct may be one means to bridge these regional and legal differences. Third party accountability agents may serve as one aspect of this oversight process with DPAs providing backstop enforcement or DPAs could provide the oversight directly.

3. Under What Authority?

Determinations that big data analytics are legal, fair and just will only be trusted and fully effective if the assessment process that demonstrates how these determinations were reached have enforcement.[17] An external body charged with this role must have authority to act in this capacity.

3a. Enforcement Authority Jurisdictional Differences

When work on the Enforcement Paper Project began, the IAF thought that the assessment framework would serve as the basis for a code of conduct for organisations engaging in big data analytics. Codes of conduct are recognised by the current EU data protection directive and by the proposed regulation.[18] In the United States, the Federal Trade Commission may enforce codes of conduct against companies over which it has jurisdiction. However, discussions with numerous DPAs revealed that while the utility of a code of conduct was evident, jurisdictional differences raise challenges to broad adoption and reliance on codes of conduct.

One of these jurisdictional differences has to do with terminology. Sometimes when a code of conduct is referenced, what really is meant is a code of practice or a best practice. Despite the difference in terminology, these different terms are used in different countries to achieve the same end—establishment and encouragement of industry best practices as it relates to maintaining comprehensive programmes. Some jurisdictions may recognise assessment processes as a code of conduct while other jurisdictions may consider them to be a best practice. In order to apply consistent terminology, from least to most formal, the instruments for best practices are:

Box 2: Instruments for Best Practices

Industry Good Practices – A shared concept, but it does not require an assertion by companies.

Suggested Industry Best Practices – A public pronouncement by a recognised

[17] Oversight and enforcement can be both internal (e.g., senior management directive) and external (e.g., regulator mandate). This paper discusses the latter.

[18] See footnote 4 supra. Laws in Columbia and Mexico also refer to codes of conduct.

authority that suggests what the authority might be looking for when inspecting industry.

Codes of Practice – For purposes of this paper, codes of practice are a statement of professional best practices by a professional organisation to differentiate from codes of conduct. (Note that some definitions use codes of conduct and codes of practice interchangeably.) Suggested industry best practices might be seen, in some instances, as a code of practice.

Codes of Conduct – The literature suggests that codes of conduct require organisations to adhere to a recognised set of rules. Those rules may come from regulations, standards groups, industry groups or even a company. From the IAF's perspective, company-based rules should link to external criteria. A code of conduct could be enforced either directly as a legal commitment or as a proxy for the underlying law that creates requirements covered by the code of conduct.[19]

Also the way DPAs currently oversee codes of conduct, codes of practice and industry good practices vary widely from jurisdiction to jurisdiction. The data protection laws of some countries (e.g., Columbia) specifically empower the DPA to oversee a code of conduct. In other countries, DPAs are not directly authorised in this way, but they may enforce the underlying legal requirements by enforcing the legal requirements the code of conduct is designed to promote. In such a jurisdiction, the law may require an organisation to balance the interests of the data controller with those of the data subject as part of establishing a legal basis to process, and the code of conduct might describe the process for doing so. In such a case, the DPA would enforce the requirement that a company engage in a balancing process rather than the methodology described in the code. In other countries (e.g., Hong Kong and Canada), DPAs serve in the role of ombudsman who may recommend industry best practices to achieve the objectives of a data protection law and may seek voluntary use of the best practices guidance. When investigating an individual complaint, these DPAs may take into consideration whether or not an organisation has implemented best practices.[20]

3b. Jurisdictional Differences When Consent is the Basis for Big Data Analytic Processing

A further differentiator between jurisdictions relates directly to the role of consent within jurisdictions. Data protection and privacy legal systems that have legitimate interests as an alternative to consent as a legal basis for big data analytic processing have a governance advantage.

[19] A code of conduct can be either an internal instrument that mandates internal ethics but is subject to external review or an external industry instrument (e.g., The Federation of European Direct Marketing Association has had an approved data protection code of conduct for years).

[20] For example, DPAs in Canada articulate guidance for achieving accountability through a comprehensive programme and use that guidance as a baseline when investigating complaints.

While there are many definitions of big data, the IAF defines big data as the use of large, diverse data sets to derive new, non-intuitive insights from the correlations between diverse data sets. Big data, by its very nature, often stretches the original purposes for which data was collected. The search for correlations is rarely the first order of business for collection or creation. Furthermore, much of the data used for big data is created – the product of earlier processing – where individuals' awareness of the data is not probable. Therefore, consent alone is not fully effective in governing big data analytics. Even if consent had been provided, effective data stewardship would require governance that goes beyond consent.

Establishing legality in many data protection regimes requires two steps: the first is the legal basis for the processing, and the second is the compatibility of the data for big data processes.

The EU data protection directive and the draft regulation establish six legal basis to process, only one of which is consent. One of the legal basis, legitimate interest, has utility where consent is not fully effective. Legitimate interest requires a balancing of the interests of the controller with those of individuals.[21] As stated earlier, the assessment process is a data protection impact assessment that, by its very nature, balances the interests of all stakeholders.

Part of oversight is establishing that the assessment process, as practiced by the organisation, meets the definition of what is required for legitimate interests. There is no question that an explicit legitimate interests regime creates greater legal certainty for all parties involved.

Most data protection regimes outside Europe require consent for processing, or satisfaction of one of a number of exemptions to the legal basis for processing. Even in parts of Europe, legal basis beyond consent is not well developed.

So how does one establish permission to process big data if consent is not fully effective and the legal basis is limited to consent? The approach would be a governance structure established by a code of conduct approximating the protections provided by the law for legal, fair and just processing. In other words, the rigor of the processing establishes the legitimacy of the processing. This is a way to bring the legitimate interest basis to countries where this legal basis does not currently exist or has not been fully recognised.

Once legal basis has been established, there is the question of compatibility. A two-phase approach to big data governance should be used.[22] The first phase is discovery, where big data is used to infer new insights based on the correlations between different data sets, and the second phase is application, where new insights are used to make real world decisions.

---

[21] Article 29 Data Protection Working Party (2014), "Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC", WP 217, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.

[22] The Centre for Information Policy Leadership (2013), "Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance", www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf.

Compatibility and legal basis need to be established in both phases. In the application phase, the organisation has the knowledge of the data to be used and the purposes for which that data will be applied, and therefore, the compatibility test is often straightforward – does the application mirror the individual's understanding of how the data will be used. In the discovery phase, a problem statement for discovering new insights may exist but a defined purpose for this discovery may not exist, and therefore, the compatibility test may be more difficult to meet because individual awareness has not been defined yet. In data protection regimes using this two-step approach for establishing the legality of data processing, establishing the legality of big data analytics is more problematic.

3c. Jurisdictional Differences When Defining Scientific Research

Where it is difficult to establish compatibility during the discovery phase, defining discovery as scientific research establishes compatibility. However, jurisdictional differences also exist when considering the definition of scientific research.

Data protection law has always respected the need to use data to create knowledge.[23] That data is often not about people.[24] For example, big data research may be about materials, the environment or supply chain processes. Data protection law only applies when data is about people.

Data pertaining to people is not necessarily personal data. Some research has integrity when the data truly is anonymised. In other cases, the data may be de-identified with both technical and policy safeguards used to protect against re-identification. Even when fully anonymised, internal policy safeguards still may be necessary, since increasingly technology alone is insufficient to protect against re-identification. If data is not personal data, then there are no restrictions on data-driven research.

However, not all research can be conducted with non-personal information. It is in instances where data can be linked to an individual that the exemption for scientific research applies.

How is scientific research defined? A session at the 35th International Conference of Data Protection and Privacy Commissioners considered the question. Some participants in the session considered research as an activity engaged in by scientists at an academic institution. Under that definition, research led by a pharmacologist at a university would be considered scientific research, but the same activity at a pharmaceutical company would not be considered scientific research.

Others participants thought that scientific research could be conducted outside a university setting but must be conducted under strict scientific methodology. Does big data research

[23] Article 13(2) EU Data Protection Directive sets forth the exemption for scientific research.
[24] Under the draft EU regulation, identification of a person may be direct, indirect or by mere fact of being singled out.

meet that definition since the scientific method requires a hypothesis and one of the attributes of big data analytics is that it often starts with a general question rather than a hypothesis?

Still, others believe research must have a public purpose linked to a public good. This definition would prohibit using a scientific research exemption for activities whose purposes are purely commercial.

It is not the role of the IAF to decide which definition of scientific research should prevail, but rather to raise the issue for consideration.[25] However, in deciding the issue, it is important to understand that separating the discovery phase from the application phase may provide safeguards for the individual to whom the data pertains. In the discovery phase, data usually is aggregated or de-identified and usually does not involve the collection of data from individuals or from the observations of individuals. Therefore, the discovery phase usually is not personally impactful. Restricting the data used for the discovery phase from the application phase, unless there are strong mitigating factors, is a way to allow further knowledge creation without creating undue risks for individuals.[26]

4. For What Purpose?

The purpose of external oversight is to provide assurance that optimised big data analytical processing achieves the objectives of the unified ethical frame, an assessment process that helps determine whether big data undertakings are legal, fair and just and helps demonstrate how that determination was reached. The EDPS Opinion makes the case that something more than mechanical checklist compliance is needed to govern advanced analytics driven by observational technologies in a universe where the cloud distorts purely national oversight. The EDPS Opinion further suggests that respect for the fundamental right of human dignity is a foundation for subsequent freedoms and rights, including the rights to privacy and to the protection of personal data. Human dignity has many components, some of which are both threatened and enhanced by big data processing. They include the ability for individuals to define themselves and not be falsely defined by predictive models. But dignity also requires the new insights that come from advanced analytics. As the EDPS Opinion argues, assurance of the full range of interests associated with dignity goes well beyond the capabilities of consent and protection requires both accountable organisations and thoughtful authorities. The governance concepts described in this paper, built on the foundation of the essential elements of accountability and supported by external oversight, are intended to accomplish the objectives suggested by the EDPS Opinion and the DPAs that have worked with IAF on the Enforcement Paper Project, data used in a legal, far and just manner enhances fundamental rights.

[25] Lack of clarity on this issue has impacted the field of data for sustainable development in parts of the world where infrastructure is limited and data from sources such as mobile phone networks would be particularly useful. For example, during the Ebola epidemic, most mobile phone companies were reluctant to provide subscribers' call records that could have helped tracked the spread of the disease due to the concerns regarding compliance with data protection laws.

[26] Abrams et al., February 2013.

To accomplish this level of assurance, the assessor has to look at all of the elements of a code of conduct (see Box 1) to see that they are in place, to determine if the organisation has the skills to put them into effect, and to ascertain that there is evidence that assessments are done with full honesty. At the end of the day, DPAs or their designate need the authority and skills to assess whether organisations have the governance and assessment tools to determine whether their big data analytic undertakings are legal, fair and just and to assess whether such governance demonstrates that assessment tools are used with integrity.

If the governance and oversight are correct, society can benefit from digital advances while still having a safe community. As discussed above, that does not mean the oversight will be done the same way everywhere. It means that the oversight mechanisms are interoperable.

5. With What Outcome?

The intended outcome is a better focus on the questions that need to be addressed to create an internal governance system within organisations that might be overseen by DPAs and on why those are the key questions. This paper also suggests various sub-topics that require continuing discussion, including:

- Recognition that an assessment process fits within a more fully developed governance structure, and it is the full governance structure that is subject to oversight.
- Understanding that big data governance processes may be characterised differently from jurisdiction to jurisdiction, but they accomplish the same ends of assuring legal, fair and just processing.

It is the IAF's intent to create a dialogue process going forward to explore these and other issues that may arise in reaction to this paper.

Conclusion

Technology should serve mankind; mankind should not serve technology. That means mankind needs the safeguards and process to harness the benefits and mitigate the risks that come from galvanizing technologies. Advanced analytics, often called big data, fits into the galvanizing technology class.

Enhancing benefits and mitigating risks related to advanced analytics is not new. Big data ethical assessments are the response to a dilemma that dates back to the 1980's when credit scoring emerged as the first big application of statistical probability when making decisions pertaining to people. In this case, it was the granting of credit to individuals. The U.S. Justice Department investigated potential discrimination related to credit scoring, and the International Conference of Data Protection Commissioners explored the same issue in Santiago de Compostela.

The revolutionary pace at which technology and business methodology evolves has only increased the necessity for understandable and trusted processes to determine whether advanced analytics in a specific instance is legal, fair and just and to demonstrate how that determination was reached. The Article 29 Working Party paper on legitimate interests[27] and the FTC paper on big data[28] both make the case that legality may rest on the question of fair and just.

The IAF, in consultation with various stakeholders, has created an assessment process that moves beyond just legal compliance to isolating and documenting the key issues in making a decision on fair and just. Actions taken based on an assessment of fair and just must be subject to oversight and enforcement to assure protection of the full range of fundamental rights and interests of stakeholders. This paper's purpose has been to spark a discussion on how interoperable oversight and enforcement might plausibly take place.

Ultimately, policymakers, such as legislators, regulators, the courts, and social consensus, will set the parameters for internal corporate processes and oversight. They will devise the legal structures that permit the data produced by an observational world to be used to create value while looking beyond privacy to respecting fundamental human rights and for regulators to interpret these legal structures in a manner that truly understands that data used wisely enhances fundamental rights.

The principal authors of the paper are Martin E. Abrams and Lynn Goldstein.

[27] Article 29 Data Protection Working Party (2014), "Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC", WP 217, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.
[28] Executive Office of the President (2014), "Big Data: Seizing Opportunities, Preserving Values", https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.