IAF Big Data Ethics Initiative
Interrogation Framework

Draft

18 March 2015

IAF Big Data Ethics Initiative
Preamble to Part B – Interrogation Framework

Introduction

Part B of the Big Data Ethics Initiative will develop a framework for interrogating big data analytics that will later be customized in Part D for specific industries, universities, companies, and public entities (the "Interrogation Framework"). Part C of the Big Data Ethics Initiative will focus on both internal and external enforcement. It is anticipated that the evidence generated by Part B may be available to internal audit and external enforcement agencies on request.

The Interrogation Framework is to be based on the previously released Part A, a "Unified Ethical Frame for Big Data Analysis" [the "UEF"] published on 7 October 2014. The UEF described a process for reaching a common ethical position on whether big data analytics is appropriate. The UEF suggests conducting a review based on data protection concepts that look beyond privacy to all the rights and interests addressed in common declarations of fundamental rights such as the United Nations Charter of Fundamental Rights. The UEF looks at the individual's rights to autonomy, but also looks at the individual's interests in health, education, opportunity and benefits from technology. The Interrogation Framework also looks at societal interests in common welfare and private enterprise's interests in innovation, competitive advantage and return on investment. In the Interrogation Framework, the individual's reticence risk[1] from not processing data is as important as the individual's interest in autonomy.

Building on prior work, the UEF separated big data analytics into two phases, "Discovery" and "Application." Generally, Discovery is where new insights, which go beyond experience and intuition, and come instead from correlations among data sets, are aggregated. Application is where these insights are put into effect and where individuals may be particularly affected in these insights are employed in an individually unique manner. The Application phase is more often individually impactful, while risks related to false insights are more of a concern in the Discovery phase. Organizations should assess the risks and benefits of analyzing data as part of both phases.

The UEF also established five key values to act as a compass for the ethics review: Beneficial, Progressive, Sustainable, Respectful, and Fair. These values are also described in more detail later in the Interrogation Framework.

---

[1] Reticence risk is the risk of not analyzing data because the data controller is unclear of whether analysis is legally or morally appropriate, and therefore forgoes a beneficial use of data.

Purpose is to Aid Judgment

Using a customized interrogation worksheet (or tool), the ethical review will identify key issues that decision makers in the organization should consider. No score will be generated that makes decisions for users. Rather, decision makers should take into account what they learn from the review process and make decisions with integrity. Documentation of the review process provides evidence that justifies the decisions made.

The judgment on whether to proceed with a concept, discovery, application and continuation of an application ultimately is based on the question of whether the analytics are fair. This judgment is based on the risks, interests, rights and benefits of all the various inside and outside stakeholders. Each of those stakeholders has a range of issues and rights. The trustworthy organization is the one that makes prudent decisions that get the balance right. For example, individuals have an interest in seclusion and autonomy, but they also have an interest in quality health outcomes and a healthier society. Sometimes conflicts between the interests may be mitigated by obscuring data elements or data created from analytics. However, there ultimately will be some conflicts that may not be resolved, and decision makers will need to decide which interests prevail and why.

At the beginning of each section of the interrogation worksheet are the key issues that should be illuminated. A read of the key questions should be conducted before doing the review.

In no way does the Big Data Ethics Initiative suggest analysis trumps privacy or data protection law. Instead, it suggests that often big data analytics does not fit into black or white governance model. As a result, an accepted process that moves a user to responsible decision making is needed.

When Should the Interrogation Framework be Used?

The Interrogation Framework should be used as big data projects reach key milestones or decision points. This will vary from sector to sector, industry to industry, and organization to organization. The Interrogation Framework should also flow naturally into the established practices of organizations.

There are four points in big data analytics where the Interrogation Framework makes sense:

Concept

Before any real analytics takes place, organizations should brain storm the reasons for using all the intended data sets, new data created, chances for new insights, usefulness of those insights, and possibilities of further application. The results of this process should be presented to decision makers for a determination on whether to proceed to the actual Discovery phase. Universities doing basic research may have concept processes that are fairly unstructured, while profit making organizations may have robust concept processes resulting in a detailed proposal before any approval can be given. The Interrogation Framework needs to work in all scenarios.

Discovery

The research to generate new insights takes place during the Discovery phase. It is

3

during this phase that data is aggregated, formatted, enhanced, or created. While this process will vary based on sector and industry, at some organizations, pre-analysis will take place at the concept phase.

The Interrogation Framework may include similar questions during both the Concept and Discovery phases. It is unnecessary to repeat questions during both phases. Rather, the interrogation should be customized for different sectors and industry practices. In some customizations, the questions will be asked in the Concept phase, and in others, they will be asked in the Discovery phase. For other customizations, the questions may need to be duplicated because of specific sensitivities. There may be several iterations that test the concept.

Application
Between the completion of the Discovery phase and the commencement of the Application phase, a decision to move forward or not is made. Beyond the objectives or interests of the organization, the organization must determine whether the analytics will create real benefits and who will receive those benefits; whether the insights will be sustainable once analytics commences; 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 does not need to be repeated. Key questions should not be ignored, and the decision maker is responsible for the integrity of the process.

Review
In order to assure controls are working, ongoing reviews are required. An ethical review should take place when routine reviews of new applications of data are scheduled. The level of the ethical review should be proportional to the constant evolution of the programs. New data sets may have been introduced, or processing shortcuts may have been developed. If changes are extensive, the ethical review should be similarly robust. However, triggers, such as complaints by individuals related to outcomes, should be taken seriously. The UEF is not about generating additional work, but creating appropriate controls, where necessary, and doing so with integrity.

Integrity is often defined as the quality of being honest and having strong moral principles. The Big Data Ethics Initiative is based on organizations weighing factors in a very honest fashion that can be trusted by all stakeholders. Examples of straying from that path might include mitigation strategies that are unlikely to work or underplaying risks to individuals related to big data analytics. Integrity is also impacted by external reviews that do not take into consideration the broad range of societal and individual benefits that might come from big data analytics. A trustworthy process requires integrity by all parties.

Key values and role in interrogation
The interrogation worksheet is based on the five key values described in the UEF and will raise

the issues necessary to make sound ethical judgments as it concerns new big data analytics. It is organized into three sections with the five values reflected throughout the three sections. The essential elements of accountability require organizations to be able to demonstrate that they have, effectively and with integrity, identified the full range of individual interests, and balanced those interests with other societal concerns. The interrogation framework and documents are intended to meet those obligations under the essential elements.

To assure familiarity with the five key values the Interrogation Framework repeats them from the Part A UEF issued 7 October 2014.

Beneficial

Both the discovery and application phases require an organization 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, organizations, 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 usually begins with a sense of purpose about the type of problem to be solved. Data scientists, along with others in an organization, should be able to define the usefulness or merit that comes from solving the problem so 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 organization should be better equipped to measure it. This application of new learnings 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 organizations may analyze data and achieve only marginal improvements but use big data because big data is new and interesting. Organizations should not create the risks associated with big data analytics if there are other processes that will accomplish the same objectives with fewer risks.[2]

---

[2] 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.

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. Models used in the mortgage securitization market to assign risk to sub-prime mortgages in the first decade of this century are examples of data scientists not understanding how the models themselves would influence the behaviour of various market players. That change in behaviour affected the model validity helping to facilitate a market decline. 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 time frame. 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.

For example, an early application of big data analytics led to a significant reduction in fraud when the discovery phase produced new insights showing a significant portion of identity fraud was not identity theft but rather came from synthetic or manufactured identities. Later insights showed that the fraudsters changed the makeup of those fake identities as organizations improved their processes to catch them. As a result, the predicative algorithms were continually refined to sustain their effectiveness in detecting and preventing fraud.

There are situations where data, particularly de-identified data, might be available for 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; and
- Canadian law allows for implied consent for evolving uses of data.

Big data analytics may affect many parties in many different ways. Those parties include individuals to whom the data pertains, organizations that originate the data, organizations 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. For example, a specialized social network might display data pertaining to individuals that they would not expect to be used for, or would be inappropriate for, employment related purposes. Organizations 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 Act prohibits unfair practices in commerce that are harmful to individuals not outweighed by countervailing benefits.[3] 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 organization 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 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, organizations should take steps to balance individual interests with integrity.

Interrogation Worksheet

The following worksheet should be considered as a model. It is not to be used as a checklist in the interrogation, but rather to be the start of sector, industry and business specific Part D documents that will be used for interrogation. Those Part D documents may be very different from sector to sector, but they should stay loyal to the objective of illuminating the issues that pertain to the five values brought forward from the Unified Ethical Frame. A reviewer should see evidence that Part D documents link to the Part B framework.

---

[3] 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.

IAF Big Data Ethics Initiative
Part B – Contextual Interrogation Worksheet

The purpose of interrogation is to identify the issues that must be resolved to assure an organization's big data project is fair to the full range of stakeholders (please see [Part A, Unified Ethical Frame](#)). The questions below have been designed to illuminate those issues for decision makers and create a record for review. For consistency purposes, it is best to stick as close as possible to these questions. However, they may be modified if different wording would be better understood within an industry as long as they still identify the key issues so fair decisions will be made.

Version 1.0, 18 March 2015

Questions	Explanatory Commentary
CHARACTERIZING THE PROJECT	
Purpose: Understand the purpose and intended outcomes of the project. Provide a project overview that describes the main purpose of the project. Is the primary purpose of the project to generate new insights or to expand on insights from a previous project or previous work?	Consider such purposes as: ▪ Marketing or risk management ▪ Building/enhancing solution and product capability ▪ Distribution network ▪ Enhancing brand experience ▪ Marketing: traditional direct mail, email, telemarketing, digital advertising, etc. (Note: Data flow mapping may be a technique that can help answer these questions.)
Sources: Understand the sources of data to be used in the project. What are all the sources of the data? Is the source data from trusted sources? What actual data elements are found in the data? How frequently should the source data be updated/refreshed? How was the data from each source originated? Can the source data be kept current over time? If not, is there an adequate replacement?	Data Origins: ▪ Provided by the individual ▪ Scraped from the web ▪ Observed in some other fashion ▪ Derived from other data ▪ Inferred from analytics Linkability: ▪ Personally Identifiable Information ▪ Pseudo-Anonymous ▪ Device Identifiable Information ▪ De-Identified ▪ Aggregate Industry obligations include codes of conduct.

Are there legal, policy, contractual, industry, or other obligations linked to the data? Is the data linkable to a particular individual or not? Is the source data structured or unstructured or both?	
Preparation: Understand the pre-processing that will be done before the analysis. What work will be done to put the data in a consistent format? How will errors and redundancy in the data be identified and dealt with? How will the data sources be consolidated for analysis? Will further synthesizing of the data be necessary?	Steps in Preparation: ▪ Data standardization ▪ Data hygiene ▪ Data integration (consolidation)
Contractual and legal conditions: All processing and applications should be within the context of the conditions associated with the data. Has there been a review of all obligations associated with the data? Is the data being used within the context of its origination? If the data is originated by others, are conditions on the data being respected? If the project moves forward, will the project security be adequate/proportional to the risks related to the data? Would the application of insights be seen as ethical and respectful if publicly exposed?	Obligations associated with the data include: ▪ Laws ▪ Policies ▪ Contracts ▪ Industry codes.
Accuracy: Evaluate the accuracy of the consolidated data. What is the accuracy of the consolidated data set to be analyzed? Are there concerns about the quality of the final data set to be analyzed?	
Insights: Understand what insights are expected from the analysis.	(NOTE: A demo can be useful in helping to understand the insights.)

What is the output from the analysis? What will the insights from the analysis be used for? Who will use the resulting insights? How long might an insight endure? What is the half-life of the insight? For how long are the insights repeatable? Can the application of the insights impact behavior in a manner that could reduce the predictive value of the insights over time? Will evolving trends impact public expectations or public policy in a manner that will impact long-term durability?	
Outcomes: Check to see that the insights and actions are progress from legacy processes. Will the project result in better outcomes than currently available? Which stakeholders have positive outcomes? Can the same or similar outcomes be achieved with fewer risks? (e.g. possibly done with less robust data)	
Accountability: Identify the individuals who are responsible for the project. Who has ultimate project ownership? Who is accountable for the various phases of the project? Do the insights contemplated by the project seem inappropriate, creepy, intrusive or rude?	Project team includes: ▪ Data capture/acquisition ▪ Data preparation/management ▪ Oversight for restrictions (legal or contractual) ▪ Appropriate application of the analysis/insights
Stakeholders: Identify all the stakeholders and their concerns. Who are all the stakeholders related to both the analysis and the use of the resulting insights? What stakeholder concerns may arise?	Possible stakeholders include: ▪ Individuals ▪ Organizations (including businesses and non-governmental organizations) ▪ Political entities/government ▪ Society/public-at-large/community ▪ Others Other factors include:

Are there other factors that should be taken into account?	- Cultural differences - Commonly held societal values - Compatibility with organizational values - Compatibility with social norms regarding the use of sensitive information.

BENEFICIAL	
Benefits: What are the benefits for each stakeholder identified above that are expected to come from the analysis?	There may be more than one benefit for a stakeholder. Obvious benefits can include: - Personalization - Health - Education - Economic opportunity - Other (please specify) - Society as whole
Risks/Mitigations: What are the risks to each stakeholder? How are the risks mitigated?	Risks to stakeholders take into account: potential impacts of false positives or negatives.
Risk/Benefit Analysis: Are the mitigated risks sufficiently balanced by the benefits? What are the residual risks after mitigation?	The risk/benefit analysis should be documented.
FAIR	
Could the result be considered unfair to individuals? If so, how? Are there Issues that could arise from this project? Will the residual risks and benefits balance individual and societal interests? From your perspective as the project owner, are you confident that the interests of stakeholders are balanced in a fair fashion?	Issues include: - Regulatory - Media - Public backlash - Discriminatory affects such as economic opportunity, physical security, physical wellbeing and limiting self-determination.