



Canadian Assessment Document (v.3)

IAF/AccessPrivacy Project

IAF Big Data Ethics Initiative

Draft

22 September 2016

NOTE: Experienced users of the assessment process may proceed directly to the assessment document

CONTENTS

[Introduction](#)

[The Five Key Values](#)

[How This Assessment Document May Be Used](#)

[Assessment Document](#)

- [PIA-Type Factors to Consider and Sample Questions](#)
- [Big Data Factors to Consider and Sample Questions](#)

INTRODUCTION

Organizations need to engage in advanced analytics to be successful today, and to do so, they need to be able to determine that their big data analytics are legal, fair and just. The purpose of this document is to introduce an assessment process that helps an organization determine whether its big data¹ activities are legal, fair and just and demonstrate how that determination was reached.

Big data is critical to innovation, and innovation is critical to a healthy, sustainable economy. The Canadian Commissioner of Competition, Innovation and Infrastructure recently remarked that: “[s]trong competition drives innovation, which in turn drives productivity, efficiency and economic growth” and that “the collection, analysis and use of data is increasingly becoming an important source of competitive advantage, driving innovation and product improvement.”² For “consumers, innovation brings more choices and higher quality products and services in a dynamic marketplace.”³

The *Personal Information Protection and Electronic Documents Act* (PIPEDA) sets out a helpful framework for a healthy, sustainable economy derived from big data analytics.⁴ PIPEDA is technology and sector neutral. By recognizing both the right of privacy of individuals with respect to their personal information and the need of organizations to collect, use or disclose

¹ “Big data can be described as data sets so large, lacking in structure, and changeable from one moment to another that traditional methods of data analysis no longer apply.” Office of the Privacy Commissioner of Canada (OPC), *Consent and privacy, A discussion paper exploring potential enhancements to consent under the Personal Information Protection and Electronic Documents Act* (“Consent Discussion Paper”), 2016, at 6.

² Remarks by John Pecman, Commissioner of Competition, Innovation and Infrastructure, in Toronto, Ontario, on May 25, 2016.

³ *Id.*

⁴ PIPEDA in its current form is equally applicable to private sector and health sector privacy law.

personal information for purposes that a reasonable person would consider appropriate in the circumstances, PIPEDA balances the interests of individuals and organizations.⁵ PIPEDA requires accountability, specifying that organizations must be responsible for the personal information under their control.⁶ However, big data analytics can create accountability challenges, and an assessment to determine whether big data activities are legal, fair and just may be part of an organization's privacy management program.⁷

Generally, under PIPEDA, organizations must obtain an individual's consent, whether express or implied, in order to collect, use or disclose an individual's personal information.⁸ PIPEDA's consent requirements establish a helpful framework for advanced data analytics by contemplating circumstances where organizations must process personal information in connection with the provision of a product or service.⁹ Specifically, under PIPEDA, organizations can require an individual to consent – within the terms and conditions of the product or service – to data analytics provided that (i) the organization complies with the specified transparency and data minimization requirements (both of which are consistent with other PIPEDA requirements)¹⁰, and (ii) the collection, use and disclosure in question is required to fulfill "legitimate purposes". Although the phrase "legitimate purposes" is not defined, it is informed by Section 5(3) of PIPEDA, which describes appropriate purposes as those that are "appropriate in the circumstances" as considered by a reasonable person,¹¹ and by the balancing of interests within the Purpose Section of PIPEDA.¹²

As a general proposition, a reasonable person would consider it to be entirely appropriate and legitimate for an organization to engage in data analytics. In order to bolster their ability to maintain their particular types of data analytics are reasonable, legitimate and appropriate in a given set of circumstances, organizations may conduct an assessment process that considers

⁵ PIPEDA's purpose is to establish rules to govern the collection, use and disclosure of information in an era in which technology increasingly facilitates the circulation and exchange of information. Section 3 of PIPEDA. PIPEDA's purpose provision, which attempts to reconcile two competing interests, privacy and organizational needs, should be interpreted with flexibility, common sense and pragmatism. *Englander v. Telus Communications Inc.*, 2004 FCA 387 at ¶ 46.

⁶ Principle 4.1 in Schedule 1 to PIPEDA.

⁷ Office of the Information and Privacy Commissioner of Alberta, Office of the Privacy Commissioner of Canada, and Office of the Information and Privacy Commissioner for British Columbia, *Getting Accountability Right with a Privacy Management Program* ("Joint Accountability Guidance") 2012, at p. 1 (Privacy management programs help "promote trust and confidence on the part of consumers, and thereby enhance competitive and reputational advantages for organizations.")

⁸ Principles 4.3.1 and 4.3.6 in Schedule 1 to PIPEDA.

⁹ Principle 4.3.3 in Schedule 1 to PIPEDA provides: "An organization shall not, as a condition of the supply of a product or service, require an individual to consent to the collection, use or disclosure of information beyond that required to fulfill the explicitly specified, and legitimate purposes."

¹⁰ The phrase "explicitly specified" is informed by PIPEDA Principle 4.2 Identifying Purposes and Principle 4.8 Openness, and the data minimization wording ("beyond that required") is informed by and consistent with Principle 4 Limiting Collection and Section 5(3) Appropriate Purposes.

¹¹ Section 5(3) of PIPEDA.

¹² Section 3 of PIPEDA.

the ethical impact of big data analytics. The Information Accountability Foundation (IAF) has developed such an assessment process. The IAF's assessment process supplements the fair information practice principles and relies on five values: Beneficial, Progressive, Sustainable, Respectful and Fair.¹³ The consideration of these five values enhances an organization's privacy management program and its compliance with its accountability obligations under PIPEDA. Ultimately, by providing a framework for establishing that the purpose and nature of a big data analytics activity are reasonable, legitimate and appropriate in a given set of circumstances, the assessment process helps an organization, as part of its privacy management program, determine whether big data activities are legal, fair and just and demonstrate how that determination was reached.

THE FIVE KEY VALUES

To understand these five values, it is important to appreciate that not all big data analytics are equally impactful on the individuals to whom the data pertains. Big data analytics can be separated into two phases: "thinking with data" and "acting with data". Generally, "thinking with data" is where new insights, which go beyond experience and intuition and come instead from correlations among data sets, are discovered. "Acting with data" is where these insights are put into effect and where individuals may be affected as these insights are employed in an individually unique manner. The "acting with data" phase often is individually impactful, the "thinking with data" phase often is not (the risks related to false insights usually are the primary concern in that phase). It often is necessary to distinguish between "thinking with data" and "acting with data" when considering the five key values: Beneficial, Progressive, Sustainable, Respectful and Fair.

Beneficial

The purpose of assessment is achieving the benefits that come with data-driven activities while mitigating the possible risks. Both the "thinking with data" and "acting with data" 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. Those risks must be counter-balanced by the benefits created for all individuals, organizations or society as a whole. This balancing concept is consistent with PIPEDA's stated purpose "to establish, in an era in which technology increasingly facilitates the circulation and exchange of information, rules to govern the collection, use and disclosure of personal information in a manner that recognizes the right of privacy of individuals with respect to their personal information and the need of organizations to collect, use or disclose personal information for purposes that a reasonable person would consider appropriate in the circumstances."¹⁴

To define benefits, one must have an understanding of why the data is being collected, used or

¹³ Unified Ethical Frame for Big Data Analysis, Information Accountability Foundation, October 7, 2014.

¹⁴ Section 3 of PIPEDA.

disclosed. 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.¹⁵ 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.¹⁶

Progressive

Since bringing large and diverse data sets together and looking for hidden insights or correlations may create some risks for some 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, then less intensive processing should be pursued.¹⁷ Precision is not required. One might not know the level of improvement in the “thinking with data” phase. Yet, by the time one is proposing “acting with data,” the organization should be better equipped to measure the level of improvement. This application of new learnings to create materially better results is what drives innovation.

Progressive must be assessed in the context in which the processing takes place. There are examples of big data being used to reduce congestion, manage disaster relief and improve medical outcomes where the level of improvement would not have been possible without big data analytics. However, there are other examples where organizations may analyze data and achieve only marginal improvements, but only use big data analytics because big data is new and interesting. If there are other methods that will accomplish the same objectives, organizations should consider pursuing those methods rather than using big data analytics to produce the same or lesser results with greater risks.¹⁸

Sustainable

Sustainable covers two issues. The first is understanding how long an insight might be effective, while the second relates to whether the data used for the insight might be available when

¹⁵ See OPC, *Expectations: A Guide for Submitting Privacy Impact Assessments to the Office of the Privacy Commissioner of Canada* (OPC Expectations), March 2011 (federal entities undertaking particularly intrusive or privacy-invasive initiatives or technologies are expected to first demonstrate that the activity or program is necessary to achieve a specific or legitimate purpose.)

¹⁶ *Id.* (The assessment should help determine whether the initiative raises privacy risks; measures, describes and quantifies these risks; and proposes solutions to eliminate privacy risks or mitigate them to an acceptable level.)

¹⁷ *Id.* (The OPC also expects federal entities to demonstrate that the intrusion on privacy is proportional to the benefit to be derived.) The concept of proportionality comes into play when conducting assessments on all of the values, but it particularly comes into play on the progressive value.

¹⁸ *Id.* (The OPC also expects federal entities to demonstrate that no other less privacy intrusive alternative would achieve the same purpose.)

acting with data.

All algorithms have an effective half-life – a period in which they effectively predict future behavior. Some are very long; others are relatively short. 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, and whether the discovery may need to be changed or refined to keep up with evolving trends and individual expectations. These considerations are consistent with PIPEDA’s Accuracy and Collection Limitation Principles.¹⁹

There are situations where data, particularly de-identified data, might be available for the “thinking with data” phase but would not be available in the “acting with data” phase because of legal or contractual restrictions.²⁰ 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. As the OPC stated in its recent discussion paper on consent and privacy, “the principle of “respect for context” bears some conceptual resemblance to the idea of “consistent use” employed in the federal Privacy Act in which a use or disclosure that is consistent with the purpose for which the information was originally collected may not require the individual’s consent. The key to employing either concept is the way in which the original “context” or original “use” is defined, since this will determine how broad a range of other uses can be considered “respectful” or “consistent.”²¹

Big data analytics may affect many parties in many different ways. Those parties include individuals to whom the data relates, organizations from whom the data originates, organizations that aggregate the data and those that might regulate the data. All of these parties have interests in the data that must be taken into consideration and respected.

An example is data scraped online for commercial purposes in violation of the sites’ terms and conditions that prohibit such site scraping. Users of big data analytics must respect those terms and conditions.

Organizations using big data analytics should understand and respect the interests of all the

¹⁹ Principle 4.4 in Schedule 1 to PIPEDA (“Information shall be collected by fair and lawful means.”); Principle 4.6.1 (“The extent to which personal information shall be accurate, complete, and up-to-date will depend upon the use of the personal information, taking into account the interests of the individual. Information shall be sufficiently accurate, complete and up-to-date to minimize the possibility that inappropriate information may be used to make a decision about the individual.”)

²⁰ *Id.* at Principle 4.4 (“information shall be collected by fair and lawful means.”)

²¹ Consent Discussion Paper at 17.

stakeholders involved in, or affected by, the analytics. Anything less would be disrespectful.

Fairness

While “respectful” speaks to the conditions related to, and the processing of, the data, “fair” relates to the insights and applications that are a product of big data. The impacts of that processing must be fair.

Canadian law prohibits discriminatory practices based on race, national or ethnic origin, colour, religion, age, and sex.²² Yet, big data analytics may predict those characteristics without actually looking for fields labeled race, national or ethnic origin, colour, religion, age, or sex. The same can be said about genotypes, particularly those related to physical characteristics. Inferring characteristics and using them to make decisions based on prohibited grounds is not fair. Big data analytics, while meeting the needs of the organization that is conducting or sponsoring the processing, must be fair to both the individuals to whom the data pertains and to whom it will be applied.²³

The analysis of fairness needs to protect against unseemly or risky actions but also to enhance beneficial opportunities. There are risks related to being too reticent with data. Human rights speak to shared benefits of technology and broader opportunities related to employment, health and safety. Pre-empting such opportunities is also a fairness issue.²⁴

In considering the value of being fair, organizations should take steps to balance individual interests in a manner that gives real weight to the interests of other stakeholders, particularly those individuals who will be impacted by the analysis.²⁵ Results should not be gamed to favor the data users.

HOW THIS ASSESSMENT DOCUMENT MAY BE USED

The purpose of this assessment document is to assist organizations to leverage the potential of big data that pertains to individuals in a manner that is consistent with Canadian law while

²² Canadian Human Rights Act, R.S.C. 1985, c. H-6.

²³ PIPEDA’s stated purpose “is to establish, in an era in which technology increasingly facilitates the circulation and exchange of information, rules to govern the collection, use and disclosure of personal information in a manner that recognizes the right of privacy of individuals with respect to their personal information and the need of organizations to collect, use or disclose personal information for purposes that a reasonable person would consider appropriate in the circumstances.” Section 3 of PIPEDA.

²⁴ Principle 4.4 in Schedule 1 to PIPEDA (“Information shall be collected by fair and lawful means.”)

²⁵ See OPC Expectations (The OPC asks government departments to answer the following four questions, which are based on the test used in *R. v. Oakes* to weigh reasonable limitations on rights and freedoms in a free and democratic society:

- Is the measure demonstrably necessary to meet a specific need?
- Is it likely to be effective in meeting that need?
- Is the loss of privacy proportional to the need?
- Is there a less privacy-invasive way of achieving the same end?)

protecting individuals from the risks of both using and not using data. This assessment document raises additional considerations that may not be covered in a typical privacy impact assessment (PIA). This assessment document does not replace PIAs; it should be used in conjunction with PIAs. It assists organizations in looking at the rights and interests impacted by data collection, use and disclosure in data-driven activities.

Organizations may incorporate this assessment document in whole or in part into their own unique processes and programs and may use a triage process to determine the questions that are appropriate to ask considering their own circumstances and the level of assessment necessary. For example, if the activity in question is only minimally changed from the past, no assessment might be necessary. If data is being used in a manner that is crystal clear from privacy notices and context, then a PIA might be all that is necessary.

This assessment document may be used as big data activities reach key milestones or decision points. Some level of assessment may be appropriate at each phase of a big data activity. Big data analytics may include phases when the activity is first conceived, then approved for programming, put into operation and eventually reviewed. Questions need not be repeated in later phases if underlying conditions have not changed. If there have been changes to the activity that impact answers, the questions may need to be repeated. If questions between the two assessments are duplicated, they may need not to be repeated, and organizations may need to determine where in their processes it makes sense for the questions to be asked.

Benefits and risks, and their likelihood, may be determined based on an organization's approach to risk. As new data analysis and new applications of insights can change over time, the process of assessing benefits and risks may need to be repeated. Regardless of when the assessment is conducted, the results of each assessment may be presented to decision makers for a determination on whether to proceed with an activity.

The assessment document identifies key issues that decision makers in organizations may consider. No score is generated that makes decisions for users. Rather, if decision makers take into account what they learn from the assessment process, decisions may be made in a manner that gives real weight to the interests of other stakeholders, particularly those individuals who will be impacted by the data analysis.

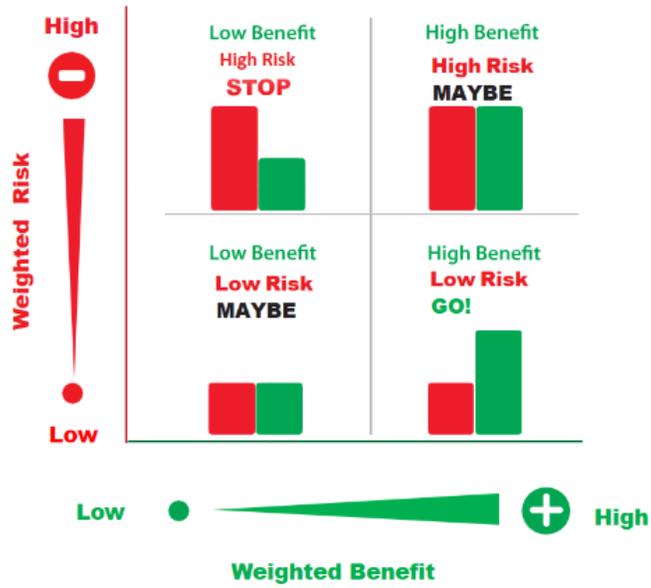
The sample questions help evaluate whether based on the assessment the activity is reasonable, appropriate and legitimate. Use of the assessment document helps determine whether the decisions reached on the appropriateness of an activity were well reasoned and demonstrate how that determination was reached. Organizations may disclose that they disclose use this assessment document or one based on it.

While developed for big data activities, this assessment document and the "thinking and acting with data" distinction may be used to assess any activities within an organization where data is collected, used and disclosed in a manner that may not have been anticipated by the individuals to whom the data pertains but is otherwise reasonable, appropriate and legitimate.

CHART

Decision – Is Acting With Data Legal, Fair and Just?

Benefit – Risk Decision Matrix – Stop or GO



Modified based on work published by the Future of Privacy Forum. For an example of assessing benefits and risk in a data intensive process, see <https://fpf.org/2014/09/11/big-data-a-benefit-and-risk-analysis/>

IAF/AccessPrivacy Project: Canadian Assessment Document (v.3)

The assessment document follows a business process flow, rather than proceeding value by value. Preceding some questions are factors to consider when answering some of the questions.

PIA-TYPE FACTORS TO CONSIDER AND SAMPLE QUESTIONS

CHARACTERIZING THE ACTIVITY

The team working on the activity should consist of people with knowledge to answer the questions in this assessment document, and the person knowledgeable about each question used should answer it.

ACCOUNTABILITY: Identify the individuals who are responsible for the activity.

Factors to Consider:

Activity team can include:

- Data capture/acquisition
- Data preparation/ management
- Oversight for restrictions (legal or contractual)
- Application of the analysis/insights

Sample Questions:

Who has the ultimate decision making authority?

Who needs to be involved in making the decision?

PURPOSE: Understand the purpose and intended outcomes of the activity.

Factors to Consider:

Examples of business needs related to the insight or the problem/question that needs to be solved include:

- Risk management
- Solution and product capability
- Distribution network
- Brand enhancement
- Marketing: traditional direct mail, email, telemarketing, digital advertising
- Service improvement
- Utilization of an existing service/product
- Education of client base
- Product development
- Market development
- Organizational effectiveness
- Process improvement
- Cost savings

Sample Questions:

What is the business need that prompted this activity?

Is this activity an expansion of a previous activity? If yes and a previous assessment was done, that assessment should be attached to provide continuity.

Does the purpose of the activity fit within a larger theme of work that is currently being contemplated or undertaken?

DATA: Understand the nature of the data.

Factors to Consider:

Examples of reasonably linkable data:

- Identifiable personal information
- Pseudonymous information
- Device identifiable information
- De-identified information

Examples of ways to anonymize data:

- Encryption

- Removal of personal information

Examples of sensitive data:

- Medical records
- Health information
- Income and tax records
- Sexual orientation
- Race or ethnic origin
- Religious beliefs
- Political affiliation or opinions

Sample Questions:

Is the data identifiable personal information?

Is the data reasonably linkable to a particular individual?

Is the data anonymous?

Are there data elements that are the product of a probability based process, such as a score?

Has the data been aggregated such that it is no longer identifiable personal information?

Has the data been combined with other data?

Is the data sensitive?

SOURCES: Understand the sources of data to be used in the activity.

Factors to Consider:

Examples of data sources:

- Individuals, collected directly by your organization
- Third parties, collected from individuals and provided to your organization
- Created through statistical analysis or calculations (identify whether this data is created by your organization or a third party)

Examples of data origins:

- Provided by the individual
- Scraped from the web
- Obtained from public sources
- Provided by third-party aggregator
- Observed in some other fashion
- Derived from data (i.e., transformation/manipulation)
- Inferred from analytics
- Provided by vendor

Unstructured data is information that is not organized in a pre-defined manner.

Source reliability/data accuracy: Data accuracy may be directly related to how data was sourced. Data directly observed may be more precise than data inferred from an algorithm. Data directly observed may be more accurate than data volunteered by individuals.

Sample Questions:

What are all the sources of the data, internal and external?

What actual data elements are found in the data?

How was the data from each source originated?

Is the source data structured or unstructured?

How reliable is the source of the data?

PREPARATION: Understand the pre-processing that will be done before the analysis.

Factors to Consider:

Organizations may have standard processes to manage data preparation. If so, this section may not be necessary once the standard process has been reviewed. It is appropriate to rely on standard processes, and it is not expected that this process will re-evaluate standard data preparation processes if they exist within the organization.

Examples of data protection:

- De-identification
- Anonymization
- Obfuscation

Sample Questions:

How will data be protected?

CONTRACTUAL AND LEGAL CONDITIONS: All processing and applications should be within the context of the conditions associated with the data.

Factors to Consider:

As a general obligation, uses of data should be open and above board. Organizations should have means of making their use of data visible to individuals, thus making consent easy to exercise. The answers to these questions relate to transparency.

Obligations associated with the data include:

- Laws
- Policies
- Contracts
- Industry obligations include codes of conduct
- Authorization
- Privacy Policy
- Privacy Statement
- Internal Policies
- Disclosures (other than the Privacy Policy and Privacy Statement)

Examples of restrictions that may apply to publicly available data:

- Website agreements
- Terms and conditions, terms of use or website policy (e.g., scraping or commercialization prohibited by website, sensor or derived data)

- Limitations on data uses from public sources

Conditions associated with data should be respected as the analysis progresses.

Sample Questions:

What obligations apply to the data?

Have all obligations associated with the data been identified?

If there are obligations, how are these obligations being satisfied?

Does consent or other legal authority exist to use the data?

Is the activity consistent with identified purposes or notices? If yes, provide the applicable wording.

Are there any restrictions on data (opt outs, withdrawals of consent, age) that would affect the use of the data? If yes, list them.

If the data has been collected by and received from others, does that party have authority and can the authority of that party be relied upon?

If the data is being shared with a third party, are there obligations that would preclude or limit that sharing?

- If there are limits on the sharing of the data, will the third party agree to the transfer subject to those limitations.

If the source of the data is public, are there any restrictions that would apply to the use of the data?

ACCURACY: Evaluate the accuracy of the consolidated data.

Sample Questions:

What steps are being taken to determine the accuracy of the data?

Is the data accurate enough for the purposes of the activity?

STAKEHOLDERS: Identify all the stakeholders and their concerns.

Factors to Consider:

Relevant stakeholders include:

- Individuals/data subjects (consumers, customers and prospects)
- Organizations (including businesses and non-governmental organisations)
- Political entities/government
- Society/public-at-large/community
- Others

Sample Questions:

Who are all relevant stakeholders related to both the analysis and the use of the resulting insights?

BIG DATA FACTORS TO CONSIDER AND SAMPLE QUESTIONS

CHARACTERIZING THE ACTIVITY

The team working on the activity should consist of people with knowledge to answer the questions in this assessment document, and the person knowledgeable about each question used should answer it.

PURPOSE: Understand the purpose and intended outcomes of the activity.

Factors to Consider:

If the initial purpose is generating insights, this assessment document may need to be repeated when the insight is applied to solve a particular question/problem or if the insight generates a particular question/problem to be solved.
(Note: Data flow mapping may be a technique that can help answer these questions.)

Sample Questions:

Is the purpose of the activity “thinking with data” (generating insights)? If yes, what insight is the activity expected to generate?

Is the purpose of the activity “acting with data” (solving a particular question/problem)? If yes, what is the particular question/problem this activity is trying to solve?

How does the purpose of the activity fit within the current business model or strategy?

How does the purpose fit within the culture of the organization?

INSIGHTS: Understand what insights are expected from the analysis.

Factors to Consider:

The questions below link to “thinking with data” (generating insights). If the original activity being assessed did not contemplate “acting with data”, all assessment questions from the beginning will need to be asked again when the insights are applied or acted on.

Insights reflect a particular point in time and can naturally change over time. The actual application of an insight may begin the process of change. Many of the questions below are designed to understand how durable those insights might be.

A demonstration can be useful in helping to understand the insights.

Insights can be a report.

Insights used beyond effectiveness can have negative impact.

Application of insights may lead to changed behaviour by individuals. This changed behaviour may impact predictiveness.

Insights may have unintended impacts. Consider additional analysis for unintended correlations. Consider how the answer relates to the scope of the activity. It will be necessary to go back to the beginning if “acting with data”.

Sample Questions:

Will the activity expand on insights from a previous activity?

How will the expected insights be used?

Are uses of the expected insights internal or external?

How long might the expected insights endure?

Could the expected insights become less useful or valuable over time?

Are the expected insights repeatable and for how long?

Could the application of the expected insights impact behaviour in a manner that could reduce the predictive value of the insights over time?

Are the expected insights reliable enough for the purposes of the activity?

Is there a less data intensive way to obtain the insights?

Is it foreseeable that the expected insights might seem inappropriate or discriminatory or might be considered offensive causing distress or humiliation?

DATA: Understand the nature of the data.

Factors to Consider:

One needs to understand the complications of mixing inferences with more objective data.

Sample Questions:

Will the same data that is available when thinking with data also be available when acting with data?

PREPARATION: Understand the pre-processing that will be done before the analysis.

Factors to Consider:

Steps in preparation of data for analysis are:

- Data standardisation
- Data hygiene
- Data integration (consolidation)

Synthesising means deriving different data elements from various source elements.

“Nature of the data” refers to quantitative data consisting of numbers (quantities) which represent counts or measurements and to qualitative data consisting of nonnumeric characteristics (such as gender or race)

Sample Questions:

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 synthesising of the data be necessary?
Will the nature of the data change when thinking with data?
When preparation is complete, will the identifiability of the data change?
What controls will be implemented to achieve data security?
Are there unique security precautions that should be applied to this particular data set?
Are there additional steps that could be taken to protect individuals without impacting the accuracy of the data?
<u>ACCURACY: Evaluate the accuracy of the consolidated data.</u>
Factors to Consider:
One cannot treat predictive values with the same confidence as verified behavioral or individually provided data. Data Transformation: Data from diverse sources in diverse formats must be put into a form where the data can be analysed. This process can impact the accuracy of the data itself and the insights that might come from big data.
Sample Questions:
Does the age of the data affect its accuracy?
Is the predictive level of inferred data taken into consideration?
Will time impact the accuracy of the data set?
Are there concerns about the accuracy of the final data set to be analysed relative to the use?

Have the sources been consolidated/transformed in such a way that accuracy is affected?
What is the potential impact of inaccurate data?
Could the impact result in discrimination?
<u>OUTPUTS: Understand what is expected from applying the insights.</u>
Factors to Consider:
<p>The questions below link to “acting with data” (solving a particular question or problem)</p> <p>Outputs reflect status at a particular point in time. Status can naturally change over time. The actual application of an insight (“acting with data”, i.e., solving a particular question or problem) may begin the process of change. Many of the questions below are designed to understand expected outputs.</p> <p>Output (application of insight) can be a report.</p>
Sample Questions:
How will the expected insights be applied?
Is it expected the insights will be applied internally or externally?
Is it expected the application of the insight will have an Impact on individuals? If yes, will the impact be positive or negative?
Is it foreseeable that application of the insight might be considered offensive causing distress or humiliation or might seem inappropriate or discriminatory?
<u>OUTCOMES: Understand the impacts on the stakeholders.</u>
Factors to Consider:

Examples of impacts to organizations include:

- Improved profitability, employee satisfaction, productivity
- Enhanced customer relationship
- Undamaged brand/reputation
- Enhanced brand/reputation
- Increased market share
- Cyber-security and fraud prevention

Examples of impacts to individuals include:

- More objective outcomes
- Safer interactions
- Better product selection
- Significant discounts
- Better product utilization
- Improved service
- Improved ease of use
- Appropriately linked to other choices, etc.

Examples of impacts to society include:

- Health care
- Education
- Climate change
- Technology

Consider evaluating expected (versus actual) outcomes.

Sample Questions:

What are the impacts, both negative and positive, to organizations?

What are the impacts, both negative and positive, to individuals?

What are the impacts, both negative and positive, to society?

STAKEHOLDERS: Identify all the stakeholders and their concerns.

Factors to Consider:

Key concerns regarding the impact on stakeholders include:

- Data use out of context
- Contract, legal, privacy or security obligations
- Data sensitivity
- Revenue/business needs
- Harm to either individuals or to the organization or to both
- Reticence risk

Other factors include:

- Cultural differences
- Commonly held societal values
- Compatibility with organizational values
- Compatibility with social norms regarding the use of sensitive information

Sample Questions:

What are the key concerns for stakeholders that may arise from this activity?

Do these concerns create material risks for the stakeholders? If yes, address them in the Risks section below.

Are there any other factors with respect to stakeholders that should be taken into account?

ASSESSMENT: Determine Activity Impact

BENEFITS

Factors to Consider:

There may be more than one benefit for a stakeholder.

Examples of benefits to stakeholders:

- Personalisation
- Health
- Education
- Economic opportunity
- Benefit to society as a whole (e.g., from producing or consuming a good or a service)
- Access to new products or services
- Consumer/customer engagement
- Brand building
- Increased revenue
- Convenience
- Price
- Other (specify)

Sample Questions:

For each stakeholder identified above, what are the benefits that are expected to come from the analysis and/or use of the insight?

RISKS

Factors to Consider:

When assessing risks to stakeholders, consider potential impacts of false positives or negatives.

Examples of risks to individuals include:

- Physical harm
- Financial harm
- Health

- Reputational
- Embarrassment
- Shock or surprise
- Inappropriate discrimination
- Misuse of data

Examples of risks to organizations include:

- Negative media attention
- Negative regulatory impact
- Compliance
- Reputational
- Business continuity
- Financial loss

Use of geo-location data raises unique risks.

Transformation of data can raise particular risks.

All processing creates risks. If using more data and robust processing does not yield more than a marginal gain, it should be avoided.

Assess whether the gain is progressive.

Sample Questions:

For each stakeholder identified above, what are the risks that are expected to come from the analysis or the use of the insights?

- For each risk, identify if it is likely to happen and what the severity is.

Can the same or similar outcomes be achieved with fewer risks (e.g. possibly done with less robust data)?

What are the downstream risks?

MITIGATION AND RESIDUAL RISKS

Factors to Consider:

Whether the risk/benefit analysis should be documented.

Risk mitigation includes outcomes.

- Evaluate expected (versus actual) outcomes.
- Consider whether false positives, false negatives, and security risks could occur.

Sample Questions:

How is each affected risk tested and mitigated?

What are the residual risks after mitigation?

Are the mitigated risks sufficiently balanced by the benefits? Explain how and quantify if possible.

IS THE ACTIVITY FAIR?

Factors to Consider:

The analysis in this section should be based, for the most part, on the answers in previous sections. Factors not considered before but considered in this section should be noted.

Issues to be considered when assessing fairness include:

- Regulatory
- Discriminatory affects such as economic opportunity, physical security, physical wellbeing and limiting self-determination
- Consistent with spirit or letter of law
- Humiliation, fear
- Reaction of media
- Reaction of public
- Reaction of family and friends
- Reaction of person accountable for activity

Sample Questions:
Are there reasonable arguments about why this activity might be considered unfair to individuals? If yes, provide details.
Are there other issues that could arise from this activity? If yes, list and evaluate them.
Are there risks to individuals from not processing data? If so, what are those risks?
Will residual risks and benefits balance individual and societal interests? If yes, explain why.
Would individuals find the activity beneficial? If yes, explain why.
Does the activity fairly treat individuals? If yes, explain how.
How will the activity impact stakeholders, either negatively or positively?

DECISION

OVERALL EVALUATION

Factors to Consider:

The analysis in this section should be based on the answers in the previous sections. Factors not considered before but considered in this section should be noted. The consideration of all these sections is to establish that the purpose and nature of the activity is reasonable, appropriate and legitimate and, therefore, legal, fair and just.

Sample Questions:

Based on the assessment, is the activity reasonable, appropriate and legitimate?