Effective Data Protection Governance Project: Improving Operational Efficiency and Regulatory Certainty in a Digital Age

Detailed Overview of Framework Components

Updated 15 November 2016

Background

The Information Accountability Foundation (IAF) was founded in 2013 to conduct research and further education on accountability-based governance. The IAF has its roots in the Global Accountability Dialog that defined the Essential Elements of Accountability in 2009. The IAF launched the Effective Data Protection Governance (EDPG) Project in 2015 to give focus to the full range of issues related to the legal, fair and just processing of information appropriate for the growing complexity of information ecosystems.

The Executive Overview paper introduces the approach and the concepts embodied in it and should be reviewed in concert with the full description of the EDPG components covered in this paper. In addition to this paper representing a complete description of the EDPG framework, it suggests approaches for implementation.[1]

The Proposed EDPG Approach and Its Components

The EDPG proposes a re-alignment of responsibilities, the introduction of new obligations and a new way to think about obligations for each participant in increasingly complex information ecosystems. The objective is to better align responsibilities while improving overall data protection effectiveness.

In recent years, there has been much debate over current governance approaches and the emphasis on data collection and purpose specification. Some have argued that a shift in focus more towards the "use" of data would provide better control and protection for the individual. The EDPG Project recognises that BOTH collection AND data use are significant and should be part of a more mature and effective governance approach. In addition, the project contemplates considering ALL data, not just data that is personally identifiable.

Moreover, while data collection and initial purpose specification are important, subsequent uses of data derived from analytical-driven insights coupled with the growing number of business participants in evolving ecosystems means a focus on data use and other factors is equally important. This proposed approach does not suggest that the current governance model, consistent with today's legal requirements, does not contemplate "data use". Rather, it suggests that this historical model, when it was developed, did not and could not have anticipated today's myriad of data-use models. Thus, a realignment of obligations is required.

The EDPG approach proposes new ways to think about individual participation (including consent), transparency and organisational accountability, while meeting the objectives of an effective privacy and data protection system to assure the fair use of information. It contemplates how and when meaningful control should be provided to individuals. It also recognises there are instances where the use of data should not directly involve the end user but, instead, where organisations should be subject to certain obligations that make sure the

[1] Peter Cullen, Jennifer Glasgow and Stan Crosley are the principle authors. They received input and editorial assistance from many project participants for which they are grateful. They remain responsible for any errors.

data and the individual are treated fairly and that data is properly protected.

Data for securing the ecosystem, operating the product or preventing fraud are examples of data use where individuals' control or choice over the use of data is not common and active engagement with individuals adds little to effective governance. These uses of data, however, should carry obligations, such as security, that should be transparent. However, they do not have to be part of a meaningful individual engagement obligation.

The EDPG approach considers a full range of factors:

- All **Participants** in an information ecosystem are considered, from individuals to all involved business entities and regulatory bodies.
- All **Data** is considered and categorised into groups that may need different treatment.
- The appropriate level of **Identifiability** for the data is considered.
- The various **Uses** of each data category are taken into account.
- The **Sensitivity** of the data itself and independently the sensitivity of the data use is considered.

Inherent in this overall approach are **five (5) interconnected components**:

1. **Ecosystem Complexity** – Effective data governance will only be successful if the approach requires ALL participants in an information ecosystem accept their designated role and corresponding responsibilities and accountability obligations.
2. **Data and Data Use Factors** – The EDPG approach recognises that both data and data use with additive considerations of "identifiability" and "sensitivity" are key to both assessing risk to all stakeholders and to developing commensurate mitigations and determining relevant obligations. More inclusive analysis of all these factors by participants will be required.
3. **Comprehensive Data Impact Assessments (Risk-Based CDIA)** – To achieve an effective mitigation of risks and realise the benefits from the opportunities to use data, a new approach to "assessments" is required. The EDPG approach builds on the concepts of PIAs but takes the assessment to a more comprehensive level assessing all relevant interests of the business and the individual.
4. **Obligations and Accountability** – New, more complex information ecosystems mean new ways to determine and action established obligations and a call for new ones. Business participants should have, in particular, meaningful, appropriate and innovative ways to engage with individuals to ensure they have suitable **participation** and a means to exercise control where relevant. Business participants should also shoulder stronger obligations to make sure fair and balanced processing of data occurs. This includes the ability to demonstrate a business participant's accountability, including internal oversight and monitoring.
5. **Enforcement** – In some jurisdictions, the shift in responsibility to business participants may mean different enforcement processes are needed to make sure data collection and use are legal fair and just. While codes of conduct may be one way to enhance enforcement and make sure business "processes" (e.g., CDIAs) are adequate, the EDPG approach recognises that regulators may need new tools to make sure effective enforcement occurs.

It is envisioned that the EDPG approach would be implemented in ways that would be compatible with local law. This approach would have the flexibility to make use of codes of

conduct or other similar mechanisms, where appropriate.

Detailed Overview of the EDPG Framework and Components—A Data- and Use-Based Approach

Component One - Ecosystem Complexity

The EDPG Project started with a particular data ecosystem in mind—the Internet of Things (IoT) where regulation is less mature and innovation is rampant—as a means to develop and test a new framework and its components. The IoT example is particularly appropriate because of the many participants, the scope of the data flows, the myriad of data uses and the rapid growth trajectory of this ecosystem. This scenario, and others, increasingly involves non-identifiable/personal data being used in ways that may have an impact on the individual.

The IoT is, in many ways, a metaphor for the challenges society is facing in implementing effective protections. To illustrate the EDPG approach, an IoT Ecosystem is used as a practical example.

The IoT is a giant network of connected "things" (which includes people and devices). The relationships can be people-people, people-things and things-things.[2] According to Cisco, by the end of this decade, there will be over 50 billion connected objects – approximately six objects for every person on the planet.[3]

The IoT ecosystem example illustrates the types of future-oriented, complex questions the EDPG approach needs to address. For example:

- What are the appropriate uses of the output of this data (e.g., could other aggregate data created through an analytical discovery process be used instead of Personally Identifiable data)?
- How should a workable framework address policy and governance guidance to leverage the same data that distinguishes between the different uses of that same data?
- While some data generated by the device may not be identifiable, it is plausible to make it identifiable by matching it with other data? What should the governance be around such uses?
- How should multiple interests be reflected and governed by society, the organisation and the individual?

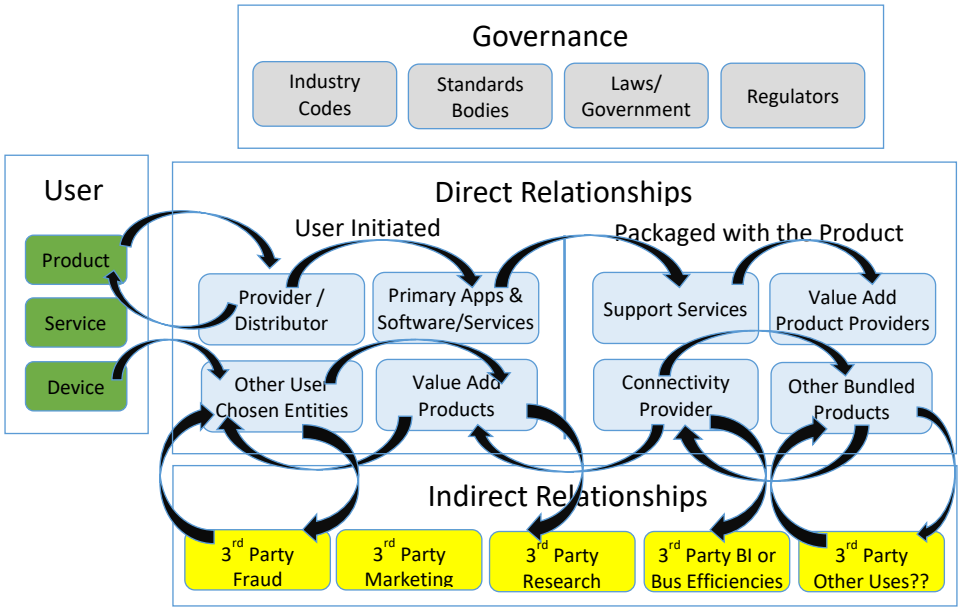
Today, most of the IoT ecosystem is less regulated in many parts of the world and does not benefit from well-established participant, industry or ecosystem information governance

[2] Morgan, Jacob, (2014) "A Simple Explanation of 'The Internet of Things", Forbes Magazine, http://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/.
[3] Endler, Michael (2013), "Cisco CEO: We're All In On Internet Of Everything", InformationWeek, http://www.informationweek.com/software/information-management/cisco-ceo-were-all-in-on-internet-of-everything/d/did/1108801.

guidance to help responsibly manage and innovatively use data produced by its products and users. Who should have access to what data and for what purposes are questions currently not answered by either laws or industry codes of conduct. These are questions that need to be addressed by the EDPG approach.

Moreover, in an IoT scenario, due to the complexity of participants, data flows and data uses, the stresses upon today's individual consent-based governance approach is particularly apparent. Also easily illustrated by the IoT scenario are the many times and places where participants, other than the end user, are responsible for many types of obligations. A 21st-century data governance approach must accommodate instances where the use of the data should not have to involve the end user, but instead, the organisation should be subject to other obligations, such as security or data retention controls**.**

Chart 1. Ecosystem Participants & Relationships



A first step in looking at the benefits and risks of an information ecosystem is to understand all the Participants. For example, an IoT ecosystem typically contains many Participants with complex relationships, including the end user (individuals or groups of individuals). Individuals and organisations in this ecosystem have complex data exchange relationships with each other. Some of these relationships are direct and exist between the User and a business Participant. Other relationships are indirect and may or may not be initiated by the User. Others are packaged with the product. Also, relationships can start, end and re-start between multiple Participants as new Uses of the product and the data it generates are envisioned. In short, the IoT ecosystem is both complex in terms of both Participants and the myriad of data flows.

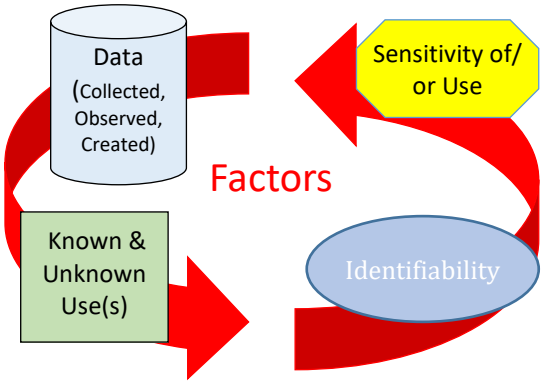
The Participant and Relationship layer identifies all the individuals, organisations (including regulators) in the IoT ecosystem who originate data or receive data from other Participants or

share data with other Participants.[4] They include the 1) Users, 2) Direct Relationships Initiated by the User or by other Participants (e.g., Direct Relationships packaged with the product, 3) Indirect Relationships, and 4) Governance entities, both regulatory and self-regulatory. A list of the types of Participants in today's ecosystems is attached as **Appendix 1.**

Component Two – Data, Use, Identifiability and Sensitivity

A key element of the EDPG Approach contemplates both data and data use with additional factors such as the identifiability and sensitivity of the data to determine the obligations appropriate for each business participant. A key part of the determination process is to fully understand each of these factors and their interactivity.

Chart 2. Key Factors for Data



Data Category Layer

The complexity of data and data flows means the same category or type of data can be used for multiple purposes, even by a single business Participant. However, each Participant may collect and use different data for different uses. Understanding the data is key.

An analysis of the data in an IoT ecosystem results in a variety of categories of data[5]. All data relating to individuals that a Participant collects, receives or generates through business processes or analytics should be classified in one of these Data Categories. Participants, however, may not have data in every Data Category. Furthermore, some Data Categories can be specific to a type of individual (e.g., customer, employee, patient).

The following is a list of the various types of Data Categories.

- **Contact Data** would include name, address, telephone number and email address.
- **Identifiers** are numerous and could range from customer identification numbers to anonymous ad identifiers.
- **Credential Data** could be the last four numbers of a User's social security number.
- **Contact Preferences** would control how the individual preferred to be contacted (e.g., email or text message).
- **Demographic Data** could include such information as age and marital status.
- **Health Data** could include such data as weight and eating habits.

[4] Any vendor or service provider, such as a cloud provider, is considered aligned with the Participant and not a recipient of shared data, unless that vendor makes independent use of the data. Vendor governance is dictated by contract with the Participant. Sharing means providing to a third party for their independent use.

[5] Included are examples of data categories. Other information ecosystems will likely require additional categories to be added.

- **Marketing/Ad Campaign Data** and History could consist of information about direct mail and/or online advertisements correlating to offers and responses.
- **Transaction Data** would include such information as order dates and amounts.
- **Sensor Preference Setting** might include "beep if the sensor battery is low".
- **Sensor Identifying Data** would include the sensor serial number.
- **Sensor Usage Data** could consist of the date and time the User started and stopped using the product.
- **Sensor Generated Data** would include device readings.
- **Geolocation Data** would include latitude and longitude history.
- **Sensor Service/Malfunction/Diagnostic Data** would include service call history and error logs.
- **Disaster Recovery Data** would include data from offsite backups.

Each of these Data Categories could be used by any number of Participants for a variety of uses in a variety of identifiable states.

Data-Use Layer

Participants collect, receive, generate and transfer data through business processes or analytics that are used for many different purposes. Data may be used internally for the Participant's own purposes, or it may be used by Users to customise their own experience. It may be used by Participants to market/sell their own products, or it may be shared with third parties for their independent Uses. The framework considers all of these Data Uses.

Below are examples of numerous Data Uses in an IoT ecosystem scenario that can be grouped as follows:

1. **Participant Internal Uses** cover the use of data to support or manage the sensor or manage the service environment, provide fulfilment to the User or manage fraud.
2. **Participant Sales/Marketing** covers use of data for marketing or advertising related to the product.
3. **Participant Research** covers the use of data to hypothesise, explore and/or identify new, different or enhanced sensors or to generate other inferences about groups of individuals for commercial purposes.
4. **Regulated Research** covers the use of data for the product Manufacturer to conduct research requiring additional legal or regulatory compliance mechanisms beyond those required by privacy and data protection.
5. **User Uses of Data (Individual) and User/Product Personalisation** covers the use of data by the User or others based on User and product instructions to tailor or change or inform the presentation and functionality of the product.
6. **Use of Data to Create New Data Products/Services (Data Monetisation)** covers the uses of data to create new data Products for use by third parties, such as Data Productisation for Fraud Purposes, Data Productisation for Marketing/Sales Purposes, Data Productisation for Research Purposes, Data Productisation for Business Intelligence**.**

A complete list and definitions of Data Use can be found in **Appendix 2**.

Identifiability Layer

Data Categories can have different and multiple Identifiability characteristics ranging from personally identified to highly aggregated. Not every Data Category will have all the various Identifiability layers.

The intent of different Identifiability levels is to recognise when different Obligations are appropriate. Altering the Identifiability level may be an appropriate risk-mitigation approach. Some Data Categories may be subject to function specific, geographic specific or industry specific laws and regulations.

For illustrative purposes, the EDPG uses four levels of Identifiability. However, the framework can accommodate fewer, more or be adapted to a different breakout.

- **Level 1** – Personal Data (PI) - Data that alone directly identifies the individual User, or through little effort can identify the User, and where administrative controls, such as contracts, allow re-identification.
- **Level 2** – Pseudonymous (PS) Data - Data that has had the direct identifiers removed or replaced with anonymous unique identifiers, such as a sequence numbers or an encrypted identifier. Level 2 data is often intended to be re-identified by the originating party but may not be re-identified by other parties who are contractually prevented from doing so. This data is sometimes known as Tagged Data or Non- Personally Identifiable Information (Non-PII).
- **Level 3** – De-Identified or Anonymous Data (DI) - Data that cannot reasonably be used to infer information about or is not easily linkable to a particular individual, computer or device[6] and for which all direct and indirect identifiers have been removed (e.g., One-Way Hashed and Salted).
- **Level 4** – Aggregate Data (AG) - Data that has had been summarised across a number of individuals and cannot reasonably be linked to a particular individual. This data is presented only in aggregated format (e.g., Census data).

Sensitivity Layer

In addition to Data Categories, Data Uses and Identifiability factors, it is important to consider the Sensitivity of both the data and/or the intended Data Use(s). Sensitivity is determined either by the Data Category, such as Biometric Data or Location Data, or is determined based on Data Use, such as for employment or credit. Once the intersection of Data Category/Data Use/Identifiability has been determined, the Sensitivity Level should be established as described below which then informs the applicable Obligations.

To add to this complexity, new uses and exchanges of Data can trigger new Obligations based on the Sensitivity of the Data, even if, by classification, the Data itself was not considered Sensitive. At the same time, if Sensitive Data is used for operational reasons, such as securing the

[6] FTC, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers,* March 2012 at 21-22, https://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers.

ecosystem, then different and even fewer Obligations on a Participant may apply.

- **Level 1 – Non-Sensitive -** Data that is not inherently Sensitive, and Data that could be inherently Sensitive, but its use is not sensitive or is permitted, for instance, by regulation (e.g., credit card data used for fraud analytical purposes, health-related data used for product operational testing).
- **Level 2 – Sensitive**
    - Data and/or Data Uses that are classified as "Sensitive" by an existing law or regulation and are required to meet the requisite Obligations (e.g., medical records, employment screening, credit limits, insurance underwriting).
    - Data Use and/or loss of data could create harm, including tangible harm (e.g., Identity theft), reputation damage, discrimination and/or targeting a vulnerable class (e.g., children or minorities).
- **Level 3 – Highly Sensitive**
    - Reserved for especially sensitive data or Data Use. (e.g., location data pertaining to visitors of a woman's shelter). A determination might inform a more active mitigation (e.g., "deletion" of data upon capture).

Component Three - Comprehensive Data Impact Assessment (CDIA)

Today's complex information flows and data uses demand the introduction of new Obligations and a new way of thinking about Obligations for each Participant in an ecosystem. These Obligations include more flexible, innovative yet meaningful ways to engage with individuals relative to their control over data about them. The EDPG approach suggests that Data/Use/Identifiability/Sensitivity all be thought about as part of a risk identification and mitigation process that establishes what Obligations are appropriate.

A cornerstone in this approach is an "assessment", scaled to correspond with the Data/Use/Identifiability/Sensitivity intersection as part of the product, service or application. Thus, the EDGP approach calls for expanding basic Privacy Impact Assessments to a Comprehensive Data Impact Assessment (CDIA), particularly for data intensive applications (product, service, analysis).

Each participant in an information ecosystem would use the assessment process as part of an effective governance approach. The framework for a CDIA would include:

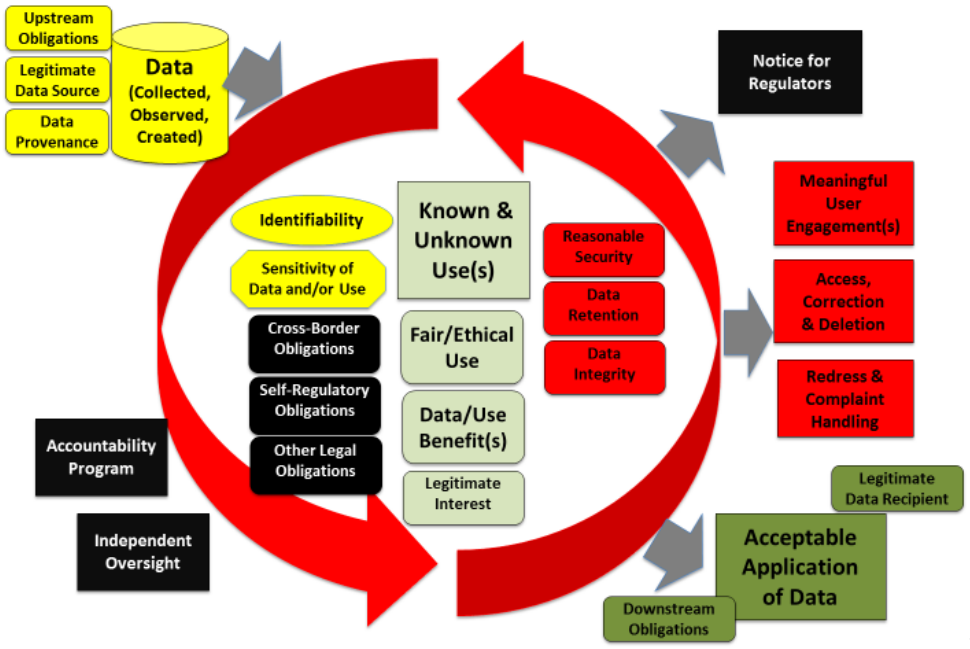
- Full project description with clear intents
- Questions to identify issues related to processing and accountability
- Clear understanding of all the stakeholders
- A description of intended benefits and possible risks
- A final assessment of fairness based on facts

The goal is to better align responsibilities while improving overall data protection effectiveness. The assessment expands typical business Participant PIAs and transparency and accountability Obligations to include an assessment that involves all stakeholders and full range of interests.

The assessment process will also be more transparent so regulators/DPAs are better informed about business practices and can be more effective in their enforcement.[7]

For any data intensive project, each participant in an information ecosystem would be accountable for assessing the full range of interests. At a high level, Data from any Data Category/Data Use/Identifiability/Sensitivity intersection may flow between any and all

Chart 3. Framework for Ethical Data Governance: Comprehensive Data Impact Assessment



Participants in an ecosystem. Obligations for the same Data Category will likely vary by Participant and Data Use.

Generally, the flow is as follows:

a. Data originates with the Participant or may be acquired from another Participant.
b. Regardless of the source, data comes with upstream obligations related to its collection or creation.
c. Participants then determine the intended Data Categories and Data Uses, determine the Identifiability factors, determine the Sensitivity factors and then determine the appropriate Obligations where mitigation of any negative impact can be maximised.

The ultimate end-user Participant of the Data may be an internal organisation or a third party. Regardless, Obligations flow downstream with the Data (when it is shared).

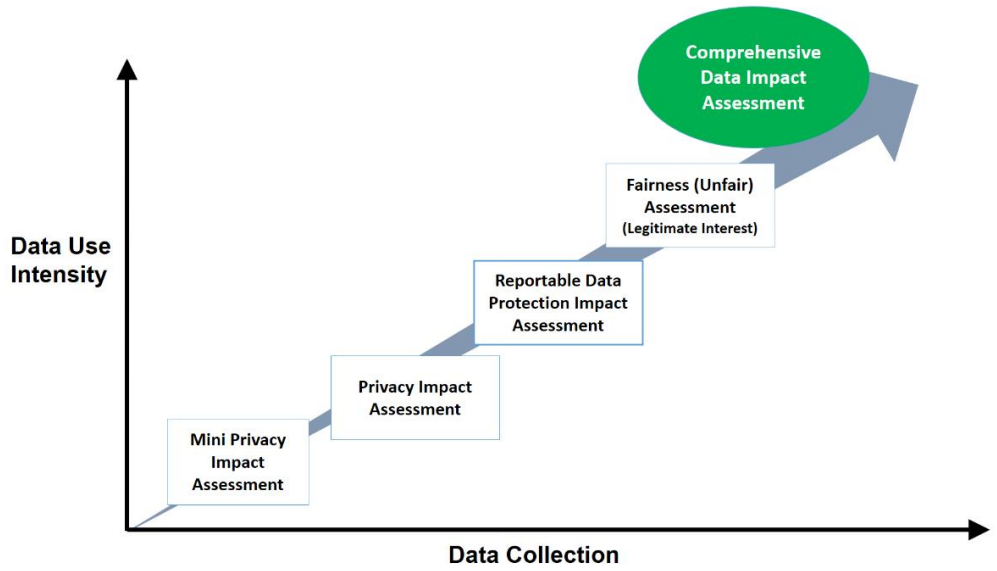
[7] Ultimately envisioned to be part of a Code of Practice/Conduct

Choosing the Right Assessment Approach

The purpose of a CDIA is to aid judgment about what obligations are appropriate for the data/use and it is particularly relevant for data intensive initiatives. Whether the project is a core product review, a new or expanded use of information, or a big data analytics project, an assessment process is required to address legal, ethical, fairness and other implications of information use including a more meaningful approach to individual participation and risk reduction. In less data intensive scenarios, a PIA maybe all that is required. Organisations may use a triage process to determine the level of assessment necessary.

The first step is to determine what type of assessment is required.

Chart 4. Assessment Choice for Effective Data Protection



The CDIA approach is intended to be customised and thus scalable, either for a Participant or for an ecosystem. A customised CDIA will identify the key issues that decision makers in the organisation should consider and ways to reduce risk to the individual. No decisions will be made by the CDIA. Rather, decision makers should take into account what they learn from the assessment and make their own decisions about what to do and not do with integrity. Documentation of the CDIA provides important evidence that supports the decision making process.

The CDIA is both linear, covering a series of definable areas to asses, as well as circular, meaning steps should be repeated to re-investigate unacceptable impacts of information use and/or identify ways to mitigate risk. This is particularly important in a big data analytics project where there is often a “discovery” (thinking with data) phase that gleans insights that may then be actionable in the “application” (acting with data) phase. It is also important to repeat the CDIA as new insights and new uses of data are identified and applied.

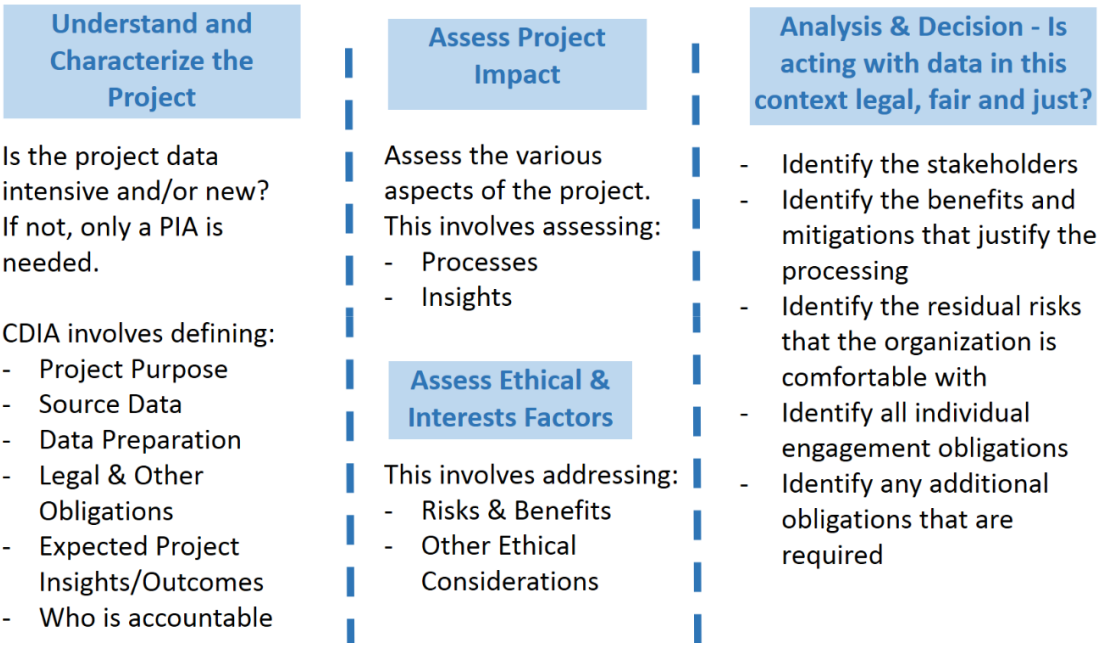
The final judgment on whether to proceed with a concept, discovery, application and continuation of an application of information insights (use) ultimately is based on the question

of whether the analytics and/or ultimate use of the data is fair and just as well as legal. This judgment is based on the risks, interests, rights and benefits of all the various internal and external stakeholders. Each of these stakeholders has a range of issues and rights. A trustworthy organisation is 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 strong economic growth, quality health treatment and an overall safer society. Sometimes, conflicts between the interests may be mitigated by a number of steps (e.g., obscuring data elements or using data created from analytics). However, there ultimately will be some conflicts that may not be easy to resolve, and business Participants will need to decide and document which interests prevail and why.

A Two-Dimensional View of a CDIA

The CDIA should be used as projects (product or big data) reach key milestones or decision points. This will vary from sector to sector, industry to industry, and organisation to organisation. The CDIA should also flow naturally into the established business and system development practices of organisations.

Chart 5. Proposed Risk-Benefits Analysis Tool for Data Intensive Initiatives



There are typically four stages in a project lifecycle and a number of logical review points where conducting a CDIA type assessment may make sense. These can be summarised as first, **Understand and Characterise the Project** followed by three assessment phases; **Assess the Project Impact** and **Assess the Ethical and Interest Aspects**. Finally, make a **Decision – Is acting with data legal, fair and just.**

It is important to review and recalibrate for maximum results. If anywhere along the way, concern is raised about any of the considerations. Or over time, as the accuracy of insights may

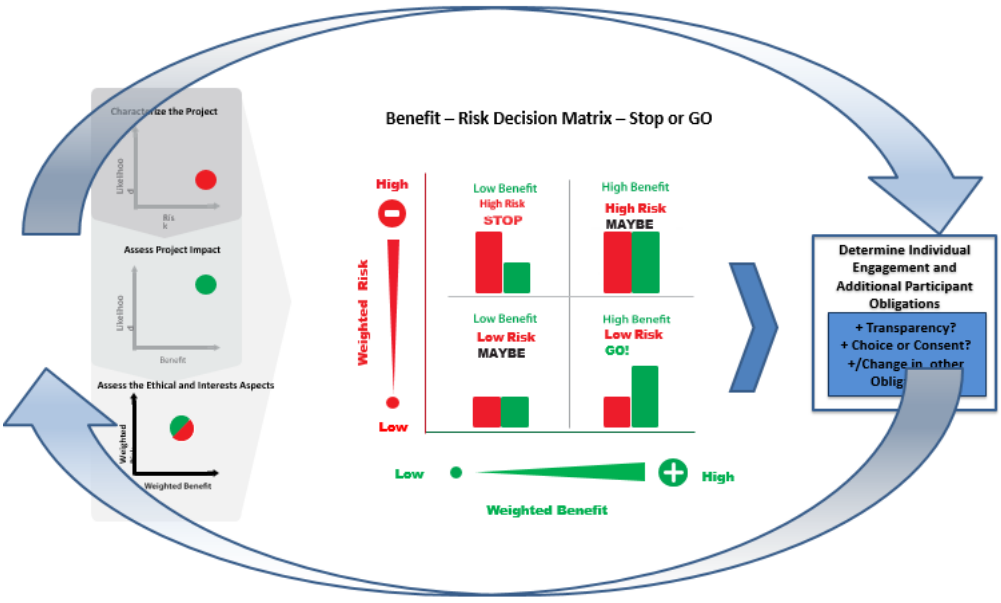
fade, adjustments in any of the factors and/or mitigations can be considered to achieve a better outcome. This includes adjustments in the data, data preparation, the actual analytical processes, as well as the application of the insights.

Recalibrating may require adjusting one or more of the factors, including identifiability, individual engagement, retention, security, accountability or any aspect of the project. In order to assure that controls, continue to be effective over time, on-going reviews are necessary. A CDIA should be done when routine reviews of new applications of data are scheduled. The level of the CDIA needed should be proportional to the evolution of the programmes. New data sets may have been introduced, or processing shortcuts may have been developed. If changes are extensive, the CDIA should be similarly robust.

In other instances, triggers, such as complaints by individuals related to outcomes, should be considered. A CDIA is not about generating additional work but is about creating appropriate controls, and doing so with integrity.

The scales for benefits, risks and likelihood should be determined based on the organisation's or industry's approach to risk (e.g., ERM programme). [8] As new data analysis and new applications

Chart 6. Decision: Is Acting with Data Legal, Fair and Just?



of insights can change over time, the process of assessing benefits, risks and changing obligations or individual engagement should be repeated (recalibrated) to achieve acceptable outcomes over time.

[8] 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/.

Regardless of when the CDIA is conducted, the results of each assessment should be presented to decision makers for a determination on whether to proceed. (See sample CDIA.) It is provided as an example with recognition organisations or industry sectors would need to modify to fit their own need, their own environments to make operational.[9]

Component Four - Obligations and Accountability Layer

The Obligations of each business Participant should be determined through the CDIA or other assessment process (e.g., PIA, DPIA, etc.).

Broadly speaking, there are two main objectives for the EDPG approach:

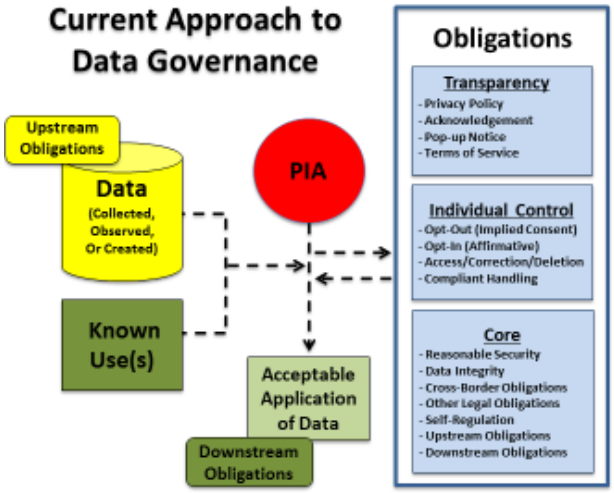
- Provide a more meaningful and relevant means of user/individual engagement and control
- Enable business to leverage information use through effective and fair information governance systems that demonstrate business governance approaches have integrity and competence to be trustworthy.

The EDPG approach expands the effectiveness of transparency by unbundling it from what regulators/DPAs need and what the Users/Individuals need in the form of User Engagement.

The EDPG Obligations can be structured under four categories:

1. Comprehensive Data Impact Assessment Obligations (assessing information risks and interests for both the User/Individual and the Business Participant - demonstrating a Business's ethical use of data)
2. Expanded Transparency Obligations through a more comprehensive notice (written for Regulators but available to Users, if Users are interested)

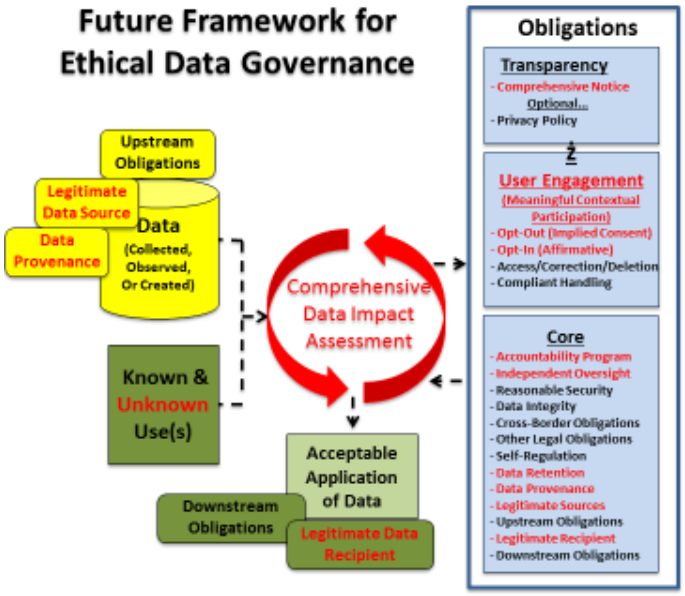
Chart 7. The Changing Approach to Governance



[9] A further example of an assessment process is being developed as part of Ethical Assessment Canadian Demonstration Project.

3. Meaningful User Engagement Obligations (where some choice actually exists)
4. Enhanced Core Obligations for the Business Participant (requiring accountability) that include some level of internal oversight.

A contrast between current obligations and a forward looking view with EDPG is shown below.



The EDPG approach also foresees a transition period where companies could use both current and the new Obligations for some period of time to bridge between the old and new approach.

For illustrative purposes "new" or "enhanced" Obligations[10] are noted in red.

EDPG Business Obligations

1. **(New) Comprehensive Information Impact Assessment (CDIA)** – A requirement that a business Participant conduct the appropriate level risk assessment considering all stakeholders and the impact for all data collection and uses (both internal and uses by third parties).
2. **Regulator and User Transparency Obligations –** Business participants would be able to choose from the transparency options listed below in appropriate combinations to achieve effective transparency for a regulator. If multiple transparency instruments are used, they may cover different activities but must be consistent.
   - **(New) More Comprehensive Notice** – This is a new type of publicly available notice that is available to Users/Individuals but not necessarily expected to be read/understood by them. Its primary audience is the Regulator. It would be a legally binding obligation in many countries. It can be as long and legalistic as it needs to be to generally describe the business practices. It would include a description of the following:
     - The Participant's practices related to the collection and use of data about an individual including self-imposed limitations on the use, sharing or other use limiting activities, or any data obscuring techniques (e.g., de-identification or pseudonymisation techniques) undertaken by the business.

[10] These Obligations have been created in the context of a Business to Consumer (B to C) environment. There may be a need to construct a variation to meet a Business to Business (B to B) and/or a Business to Business to Consumer (B to B to C) environment.

- The Participants practices related to the Standard Obligations including a general description of the types of oversight and monitoring done by the business and what the Participant does when a new use of data occurs that affects the individual and triggers some form of User Engagement as described below.

- **Traditional Privacy Policy** – Current privacy notices (both long and short form) written for both consumers and regulators. These are usually in a relatively consumer friendly format and may reference the More Comprehensive Notice.

3. **(New) User Engagement Mechanisms to Achieve Informed Choice** – The concept of User Engagement encourages multiple innovative approaches to what has historically been some type of informed choice about a specific use of data. User Engagement activities are intended to do the following:
   - **Meaningful Choice** – User Engagement provides a wide variety of scenarios and mechanisms where meaningful choice is either required or is appropriate.
   - **Appropriate Access, Correction and Deletion –** User Engagement may also include the ability for the User to see appropriate Data about them held by a business Participant and to change or delete the Data, if applicable, or suppress its Use.
   - **Proper Timing** – User Engagement would offer controls/choices that are provided at an appropriate time for a new use of data that involves some choice by the User/Individual. This can occur when the Users'/Individuals' interests need to be explained (e.g., context based Pop-Up notices like a request to use location data). The EDPG approach encourages new innovative choice mechanisms not currently invented be developed.
   - **Context Driven –** User Engagement is driven by the context of the Data/Use/Identifiability/Sensitivity intersection and where a User is meaningfully engaged. User Engagement is not intended for expected or commonly understood collection or Uses of Data, or where there is no meaningful choice (e.g., Fraud management).
   - **(New)** User Engagement options offer at least two levels of control for the Individual.
     - **Level 1 – Opt-in or Affirmative Consent**
     - **Level 2 – Opt-Out or Implied Consent**
   - **Redress & Complaint Handling** – The ability for a User to engage with any Participant to ask question, receive clarification or resolve a problem they believe they have with the business and how data is handled.
4. **Standard Obligations –** Standard bedrock Obligations that every business Participant would adopt to an appropriate level.
   - **Reasonable Security** – Protect data appropriate for the Data/Use/Identifiability/Sensitivity.
   - **Data Integrity** – Be sure data has a sufficient degree of accuracy for the intended Use.
   - **Incident Response –** Manage security and other privacy incidents.
   - **Cross Border Obligations** – Comply with all the obligations that come as a result of moving data across borders.

- **Other Legal Obligations** – Comply with any applicable obligations that are specifically codified into law.
- **Codes of Conduct** – Voluntary compliance with Obligations that a Participant has agreed to from a specific industry or sector code of conduct or practice.
- **(New) Data Retention Practices** – Appropriate policies and practices regarding retention of data, keeping it only as long as needed for the business purpose.
- **(New) Accountability Programme** – A proportional programme requiring a business commitment to internal policies, mechanisms to put polices into effect and monitoring to make sure mechanisms work.
- **(New) Independent Oversight** – Some form of independent assessment of the Accountability Programme, including the CDIA process that could be performed by an internal audit group, an outside party (e.g., Assurance Assessment) or an IRB, through an industry code of conduct, a certification process (e.g., PCI standard) or another similar independent oversight mechanism.
- **(New) Legitimate Source** – An assurance process beyond contract warranties to make sure 3rd parties from whom data is obtained are legitimate entities and have legitimate data collection practices.
- **(New) Data Provenance –** Understanding the provenance of the Data (where it is from, the quality, the recency, the accuracy) as part of the CDIA.
- **Upstream 3rd-Party Contract Obligations** – Understanding and complying with all policy and contractual obligations that come from internal sources or from 3rd parties providing Data.
- **(New) Legitimate Recipient** – Taking steps beyond contract warranties to make sure 3rd parties to whom Data is shared are legitimate entities and will honour the terms of the data transfer/license agreement.
- **Downstream 3rd-Party Contract Obligations** – Passing on all obligations that should be part of the transfer agreement or contract when Data is shared with another internal entity or a 3rd party.

The EDPG approach incorporates the work done in recent years by the IAF on the Essential Elements of Accountability and Big Data Analytics. Accountability requires that organisations demonstrate that they have, effectively and with integrity, identified the full range of individual interests and balanced those interests with other societal concerns. The CDIA adopts Accountability as one of its key components.

The EDPG approach also incorporates the five key values identified in the IAF work on big data analytics. These values are important for any application of data that may have an impact on an individual, but are particularly important when the project involves extensive data/big data analysis. (See **Appendix 3** for a full description of values.)

Component Five – Enforcement

Subsequent versions of this document will cover the final component of the EDPG approach, specifically the **Role and Function of Enforcement. In addition, as a key function of both accountability and enforcement is the "demonstration" of accountability, subsequent versions will also address this area.**

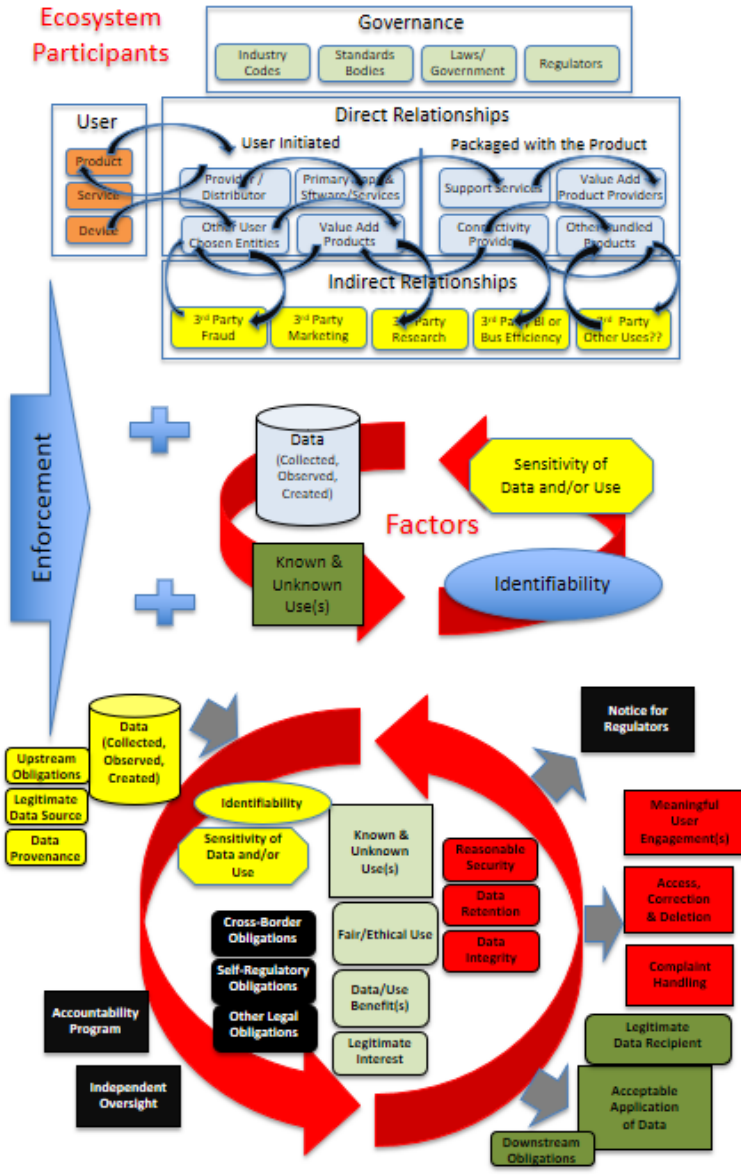
Summary and Benefits of the EDPG Approach?

Today's privacy and data protection systems are often dependent on the individual to make choices about with whom they will share data, make decisions on reasonable uses of that data, and make complaints to regulators/DPAs when data are used beyond the bargain. While maintaining the spirit of these individual involvement goals, the EDPG approach provides a fundamental realignment of responsibility for assuring more effective data protection from the individual to the organisation.

The approach is a set of **five (5) interconnected components:**

1. **Ecosystem Complexity** –ALL participants in an information ecosystem must accept their designated role and corresponding responsibilities and accountability obligations.
2. **Data and Data Use Factors** – Both data and data use with additive considerations of "identifiability" and "sensitivity" are key to both assessing risk to all stakeholders and determining relevant obligations.
3. **Comprehensive Data Impact Assessments (Risk based CDIA)** – A new approach to "assessments" is required: a more comprehensive level of assessing all relevant interests of the business and the individual.
4. **Obligations and Accountability** – New, more complex information ecosystems mean new ways to determine and action established obligations and create new ones. Individuals should have suitable **participation** and a means to exercise control where relevant. Business participants will also shoulder

Chart 7: Effective (Accountable)Data Protection Governance



stronger obligations to accomplish fair and balanced processing of data.

5. **Enforcement** – Different enforcement processes are needed to ensure data collection and use are legal fair and just. While Codes of Conduct may be one way to enhance enforcement and make sure business "processes" (e.g., CDIAs) are adequate, the EDPG approach recognises that regulators may need new tools to achieve effective enforcement.

The EDPG approach includes:

- **Individual engagement** in data collection and data use issues where they can participate in a more easily understandable way. The approach therefore does not rely on notices to inform individuals, and it does not expect individuals to govern through those unread and unreadable notices. Instead, the approach provides individuals with more meaningful and flexible engagement in situations where there is some impact to the individual and where individual control is effective. If an individual is interested, there is still the longer, more legalistic notice written for regulators/DPAs available to them, but they are not expected to read it. Individuals are able to focus on information practices of organisations where they actually have a meaningful opportunity to participate.

- **Organisations will conduct CDIAs**[11] based on guidance from implementation of the EDPG approach through industry codes or other similar approaches appropriate to the jurisdiction that identify the type of user engagement required and the appropriate obligations for each business Participant. The obligations often vary for each Participant based on their roles in the ecosystem. The CDIA process considers the Data Categories, the Data Uses and other mitigating factors, including Identifiability and Sensitivity. CDIAs look at all immediate implications to individuals and also consider the possible implications over time. Adverse implications for individuals are addressed with appropriate mitigations (e.g., de-identification of the data, extra security).

- Organisations have the freedom to conduct CDIAs to determine their transparency and individual participation obligations but will have to **demonstrate to regulators/DPAs**, if asked, that these assessments are fair to the individual. Where appropriate, codes of conduct can be used to establish guidance for organisations in assessing what is fair to the individual and clarity for an organisation's ability to leverage data for knowledge creation.[12]

- Organisations will **provide greater transparency** in a notice that will also cover their policies and practices relating to information collection and use as well as information governance (including its CDIA process).[13] This notice is likely more robust than many privacy policies today and constitutes the obligations a regulator/DPA looks to regarding compliance oversight. While this disclosure is accessible to both individuals and regulators/DPAs, it is

[11] A full CDIA is contemplated for a data scenario that is intensive. A scaled down version would be used for a less data intensive scenario encompassing the goals of a PIA/DPIA.
[12] A Code of Conduct or Practice may build upon the Big Data work involving enforcement - http://informationaccountability.org/iaf-workshop-examines-big-data-regulation-and-enforcement/.
[13] Content parameters of this detailed type of notice are still in development. It is anticipated the overview of an organisation's governance would be high-level with more information available to a regulator/DPA upon request.

written for regulators/DPAs. Individuals are not required, or expected, to read or acknowledge it.

The EDPG approach recognises that privacy, the protection of individual autonomy, data protection, and the assurance of fair use of information, are important aspects of protecting values and interests in diverse legal systems.

Its benefits include, more effective ways for individuals to engage with the use of information about them and more contextual ways to participate in meaningful control over that data are expected to emerge. In addition, with more accountability for organisations to responsibly use information about individuals, there is less risk to the individual, and there will be higher confidence that organisations are focused on addressing a complete set of individual interests associated with the use of information about them.

The approach would be implemented in a manner that builds on the way local laws work. In the longer term, in some jurisdictions, one might see codes of conduct that are directly or indirectly enforced by authorities. In other jurisdictions, accountability agents might have a role.

The EDPG approach creates more accountability/enforceability over responsible business use of information, including areas that are not subject to direct regulation and/or where a direct individual relationship may not exist. The approach will enable a more thorough application of privacy principles by organisations that include what is currently non-personal data and therefore is not covered by most data protection laws. In addition, it provides more effective information governance covering a broader range of interests, including more meaningful and innovative ways in which individuals are engaged.

It eases the burden on regulators/DPAs by providing more transparency about how organisations collect and handle data. A comprehensive notice written by lawyers for regulators/DPAs should help them be more effective in their enforcement activities. In addition, the regulator/DPA will have a more direct route directly to the organisation to enforce the law.

Finally, the EDPG approach creates greater confidence that organisations' information governance systems are meeting a broader set of expectations. It facilitates managing a broader set of interests and risks that allows organisations to leverage information to create value. Through codes of conduct or similar processes, organisations will have greater clarity about what is expected of them, including regulator/DPA expectations for an accountable process to establish objective based obligations.

Next Steps

It is the intent of the EDPG Project to expand the focus beyond the IoT ecosystem, testing it against other business models. The EDPG Project is also intended to evolve and change over time as technologies and business processes change.

The IAF has been working internally on the EDPG Project since early 2015, and while parts of the EDPG Project are more advanced and are ready to socialise, test and refine with other stakeholders, other parts are less developed. For example, the framework component parts

addressing Accountability and by extension Oversight and Enforcement and a model to address types of data impact areas that should result in additional individual engagement, are areas where more development is planned.

Over the coming months, the IAF plans on further developing, testing and socialising this approach, including exploring how the entire approach and components may fit into or support the implementation of local laws. For example, the adoption of the European GDPR, with its elements of a risk-based approach utilising enhanced accountability, is the beginning of a global change process to assure modern information processes are governed in a fashion that protects individuals while facilitating digital economies.

The testing, socialising and further development of the EDPG approach will be accomplished through dialog with multiple stakeholders who all share the same goals of enabling the generation of opportunities and benefits from information while effectively protecting individuals and considering the broad range of interests relative to its use.

APPENDIX 1: Participants and Relationships

(Back to text.)

ECOSYSTEM PARTICIPANTS

There are 3 types of active Participants described below and a 4th who establishes and influences Data governance obligations, but may also be involved in receiving/transferring Data. Data may flow back and forth between any of the participants for a given situation, and obligations may be different for each Data flow.

- Participant(s) means
    - any natural person, legal entity, public authority, agency or other body, or any Product that
        - originates [generates?] Data,
        - receives Data from other entities, or
        - Shares Data with other entities
    - and which utilises the Data for their own/independent purpose.

- Sharing [disclosure?] -- a disclosure of Data by transmission or dissemination, or otherwise actively or passively making Data available to a Participant, for that Participant's independent use.

- "Vendor" means a recipient of Data acting on behalf of a Participant that only receives data from other parties for the use and processing purposes of the Participant with which they are contracted and not for the recipient's own/independent use. Vendor governance would be dictated by contract with the Participant. If a Vendor has the right to use Data for their own independent purpose, the Vendor would be considered a Participant with respect to that independent use.

PARTICIPANTS

Users and Products

- User(s) – a natural person who directly interacts with a Product
- Product(s) – includes products, services, or devices that originate, receive, or share Data and are being accessed or used by the User (e.g., mobile application, call centre service, cable TV service, health care or smart shirt).

Direct Relationships with User

- Entities with a User Relationship Initiated by the User
    - Product Provider, Manufacturer or Developer (e.g., device manufacturer, mobile app developer, website developer, social media site or call centre)
    - Product Distributor or Delivery (e.g., application webstore, retailer, distribution and shipping company)
    - Other User-Chosen Participants (e.g., pharmacy, insurance, care-givers, legal guardians)
    - Complementary Product Providers – entities that provide Products that are complementary to, but not required for, the use of the primary Product (e.g., universal remote, digital pill minder)

- Derivative Relationships with User – Providers whose direct relationship with User arises due to the Product's [required?] [dependent?] use of the secondary Product (such as products packaged or bundled with the primary product)
    - Support Service Providers (e.g., Apple support, maintenance services)
    - Value-Add Product Providers (e.g., phone service, warranty, insurance)
    - Connectivity Providers (telecommunications providers, wireless providers)
    - Other Bundled Product Providers

Third Party Relationships -- entities who receive Data but have no direct relationship with the User and who are not Vendors [NOTE: "3rd Party" introduces a term that we should use consistently or perhaps replace with "Indirect"]

- 3rd-Party Fraud Service Providers
- 3rd-Party Marketing Service and Data Providers
- 3rd-Party Research Service Providers
- 3rd-Party BI or Business Efficiency Providers
- 3rd-Party Other Uses *(to be defined)*

Governance/Oversight/Enforcement (Set/Influence the obligations of a Data Governance system)

- Industry organisation
- Standards Bodies
- Courts/Litigants
- Regulators

APPENDIX 2: Data Uses

(Back to text.)

PARTICIPANT INTERNAL USES

Operations

- Internal Operations – Use of Data by a Participant to support the activities, decisions and responsibilities of managing the resources which support the production and delivery of Products by the Participant. This issue includes security.
- Business Improvement – Use of Data by a Participant to improve their own business operations. These activities may not be directly related to a Product.
- Ecosystem Operations Management – Use of Data by a Participant to maintenance and improve the supporting ecosystem (e.g., Ecosystem Security, Data Load Balance, etc.).

Order Fulfilment and Support

- Order Fulfilment – Use of Data to initiate and conclude the delivery and service of the Product.
- Support Services – Use of Data to provide direct and indirect support that ultimately supports all functionality of the Product, including hosting and storing, running analytics to inform and perhaps maximise value of operation of the apps and devices, troubleshooting, efficiency analysis, user help desks, etc.

Product Management

- Product Development – Use of Data to develop and deliver new products or services, including data related products, with the intent of commercially marketing the resulting Product. This includes Product cycle of design, develop, test, and release. This includes applying the results of Research/Discovery to a specific Product.
- Product Improvement – Use of Data to identify and make upgrades or improvements to the Product, including, for example, next generations of the product, and/or Use of Data to manage the lifecycle of the Product.
- Quality/Safety – Use of Data to support, report and/or assess quality and safety issues related to the Product, and/or related components (e.g., apps or other services).

Fraud Detection/Prevention

- Fraud – Use of data for detecting and/or preventing all types of fraudulent activities.

Legal/Regulatory

- Legal/Regulatory - Use or retention of any Data to support efforts by a Participant to comply with law, defend its intellectual property or its business operations, or to satisfy legal or regulatory requirements to enable Participant to conduct its business. This includes the request for Data pertaining to an individual or a group of individuals, formal (e.g., with a subpoena) or informal (e.g., a “request for information”), by Government and law enforcement agencies. These uses are often beyond the control of the Participant that has Data in its possession or control.

Public Purpose

- Public Purpose –Use or repurposing of Data for "public good" or standard business reporting for public purposes. Traditional concepts of "public good" include providing infrastructure services (e.g., roads and utilities), shelter, food and safety for individuals, to stop the spread of deadly disease, avert natural catastrophe, economic forecasting, or to provide aid in the form of disaster relief. Standard Business Reporting includes periodic reporting of financial results and general industry and geopolitical economic forecasting and modelling.

PARTICIPANT SALES/MARKETING

- Marketing/Advertising – Use of Data to market and advertise new Products and Products related to the primary Product. This spans both offline and digital marketing (e.g., online, mobile and addressable TV).

- Sales – Use of Data to sell new Products and Products related to the primary Product.

PARTICIPANT RESEARCH

- General Commercial Research/Discovery – Use of Data to hypothesise, explore and/or identify new, different or enhanced Products, or to generate other inferences about groups of individuals or Products for commercial purposes. This category includes the ideas related to more "pure" research where the goal may not be known in advance.

- Public Purpose Research/Discovery – Use of Data to hypothesise, explore and/or identify new, different or additional Products, or to generate other inferences about groups of individuals or Products for a public purpose.

PARTICIPANT REGULATED RESEARCH

- Legally Restricted Research – Refers to data-based research, broadly defined, that requires additional legal or regulatory compliance mechanisms beyond those required by privacy and data protection in order to be conducted by the Entity (e.g., Human Subject Medical Research (as defined by regulation)

USER USES OF DATA (Consumer)

- User/Product Personalisation – Use of Data by the User, the Product or others based on User and Product instructions to tailor or change or inform the presentation and functionality of the Product. This may be done by instructions or choices from the user or by serving usage of the Product and adjusting presentation and functionality contemporaneously.

USE OF DATA TO CREATE NEW DATA PRODUCTS/SERVICES (DATA MONITISATION) –

The use of data to create a new data Products for use by 3rd parties.

- Data Productisation for Fraud Purposes – Developing Product from Data by the Participant for fraud purposes of other Participants. May or may not be for a fee.

- Data Productisation for Marketing/Sales Purposes – Developing Products from Data by the Participant for marketing purposes of other Participants. May or may not be for a fee.

- Data Productisation for Research Purposes – Developing Products from Data by the Participant entity for research purposes of other Participants. May or may not be for a fee.
- Data Productisation for Business Intelligence – Developing Products from Data by the Participant for BI for other internal efficiency purposes. May or may not be for a fee.
- Other Regulated Data Productisation – Developing Product from Data by the Participant for regulated purposes of other Participants (e.g., credit, insurance, employment, etc.).

APPENDIX 3: Values

(Back to text.)

What are the Values Included in the EDPG Approach?

Benefits and Risks – Both the discovery and application phases of a big data analytic project require that an organisation define the benefits that will be created by the information use and should identify the parties that gain tangible value from the effort. The act of information use, including 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. While data/big data analysis 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 organisation, 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 result 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 Application – Because the mere process of bringing large and diverse data sets together and looking for hidden insights or correlations may create risk for individuals, the value from data/big data analytics should be materially better than not using this information. If the anticipated improvements can be achieved in a less data-intensive manner, that less intensive processing should be pursued as a form of data minimisation. 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. The application of new insights to create materially better results is often referred to as innovation. Organisations should not create the risks associated with data/big data analytics if there are other processes that will accomplish the same objectives with fewer risks.

Sustainable Application – A key question is whether the benefits are sustainable? For example, 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. Data/big data analysts should articulate their best understanding of how long an insight might be used. Data/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 continues to have 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 the discovery phase but would not be available in the application phase because of legal or contractual restrictions.

Respectful Application – Respectful relates directly to the context in which the data originated and to any 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, 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. For example, a specialised social network might display data pertaining to individuals that they would not expect to be used in that way, or would be inappropriate for, employment related purposes. 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.