Information Accountability Foundation

Making Data Driven Innovation Work

Lynn Goldstein and Peter Cullen

February 2023

Background and Project Overview[1]

The use of data to create knowledge in an academic setting is well understood. Less appreciated is the use of data in the commercial setting to create knowledge. Increasingly, digital agenda in both the public and private sectors are about creating safe pathways for data to be turned into information, for information to be turned into knowledge and for knowledge to facilitate actions that are societally beneficial to people and for this processing to be conducted in a manner that is lawful, fair, and just.

Knowledge creation in digital ecosystems is the generation of new insights about people that originates with advanced analytics (such as artificial intelligence) on data collected directly, passively, or observationally. The data is not about a specific individual, and, at this stage, it has a significantly lower impact on an individual. Knowledge creation also can be called "thinking with data." It is key for processes such as data analytics used in product or service development and improvement. It may more broadly be called "data-driven or digital innovation."

Knowledge creation, which is the pathway to the future, is conducted for three different research purposes: (1) scientific, (2) commercial, and (3) hybrid (public-private collaboration). When used as part of advanced data analytics, knowledge creation is the engine that drives the digital economy in a society.

Knowledge creation differs from knowledge application. Knowledge application utilizes knowledge created on a specific or identifiable set of individuals. The process of "knowledge application" has been called "acting with data." Understanding the differences between knowledge creation and knowledge application, and the purposes for which they are being undertaken as well as the different set of risks for each, is critical to understanding how they should be regulated.

The Information Accountability Foundation (IAF) believes it is critical that organizations be able to think with data and to engage in knowledge discovery and creation in order to achieve a trusted global digital ecosystem. The IA has advocated for many years that there should be a distinction between knowledge creation (thinking with knowledge) and knowledge application (acting with knowledge). Increasingly, this sort of data processing leads to the creation of new insights key to digital innovation. However, the current pattern of data protection law development, and the enforcement of these laws, does not appreciate the distinction between these two processes, and this failure has led to confusion and hesitancy about when advanced analytics for knowledge creation may be used where there is no distinct legal basis for processing personal data for this purpose.

Knowledge creation in and of itself is less impactful on individuals. Knowledge application leads to decisions and actions that impact individuals and brings into consideration traditional privacy concerns. Knowledge creation does not have the same concerns and results. Knowledge creation and knowledge

[1] This paper was prepared by Peter Cullen and Lynn Goldstein with edits by Barbara Lawler and Martin Abrams. The opinions in this paper reflect their views, and are not necessarily the views of the IAF Board, its board members, or other members of the IAF community.

application should not be treated the same way. Therefore, regulatory approaches, including regulation, oversight, and enforcement, should treat these two processes differently. Both processes require controls, but since the risks are not the same in knowledge creation and knowledge application, the controls should not be the same.

Treating knowledge creation and knowledge application the same can have the unintended consequence of overburdening the organizations that are the engines for digital innovation (e.g., innovators in medical devices, transportation, education, design, and services.) There is a great deal of concern about surveillance market players, but they are not the only knowledge creation innovators.

Furthermore, at its extreme, polarization is affecting the public policy debate on privacy and data protection. Terms such as autonomy and control increasingly are being interpreted as personal sovereignty. Policymakers designing the new data-driven industrial policy are talking past independent regulators who are concerned markets are dominated by data extraction that harms all individuals, particularly protected classes. This situation is making legislation and regulation that both facilitates innovation and protects the full range of stakeholder interests increasingly difficult.[2] This void has led some organizations to delay or forgo the creation of these insights. As a result, the trajectory of the application of data protection public policy has the potential to stifle further data driven innovation.

The IAF undertook this research to understand how organizations discover and create new knowledge. This project seeks to clarify the impact of regulatory and public policy uncertainty on commercial-driven knowledge creation, develop scenario driven examples of this impact, develop a public policy model that enables responsible data-driven knowledge creation through a series of compensatory controls and create a narrative and path for knowledge creation to be more formally recognized as legitimate data processing activities in next generation privacy and data protection law[3].

Problem Introduction

Many organizations use personal data as part of analytics processing (Corporate Research) to solve identified business problems, but most organizations do not use Corporate Research as a distinct processing activity more broadly because:

- Most organizations generally do not break data processing into two distinct phases: knowledge creation (I.e., the research function to identify a solution to a business problem) and knowledge application (I.e., application of the solution to the business problem), and/or
- In the EU, this portion of data processing (research) is complicated and/or limiting. Personal data can be processed for scientific research purposes, which is narrowly defined, as long as sufficient safeguards have been implemented and scientific research is run “in accordance with relevant sector-related methodological and ethical standards, in conformity with good practice” (collectively Safeguards), and/or

[2] See detailed overview of Public Policy polarization in Appendix II

The IAF research team conducted extensive interviews with organizations from numerous fields that use data to create knowledge, even if they did not identify their processes in that way. The IAF team used the interviews to supplement the team’s decade of work as researchers at the IAF and consultants working on ethical assessments and demonstrable accountability.

- Other countries and future laws limit the use of personal data for research purposes to improvement/supply of products and services or to requiring the use of de-identified personal data for research and for socially beneficial purposes.

The IAF thinks that organizations might be able to use personal data more broadly for Corporate Research if, in addition to legal and regulatory modifications, appropriate Safeguards were put into place and proposes that those Safeguards be developed and implemented in a defined "Research Sandbox" environment.

Circular Reinforcing Problem

The IAF has long posited that there is a processing and, by extension, a risk distinction between data driven knowledge creation and knowledge application. Overlaying this distinction is a "nomenclature" challenge; many organizations generally do not use the term "research" as a function, and the terms "knowledge creation" and "knowledge application" do not fit their current business processes. These organizations more commonly use the term "data science." In AI, the term "experimentation" is more common and applies to a specific stage of model development. For the purposes of this project, the term "Corporate Research" will be used to refer to phases of data driven analysis (AI or other methodology) that are the precursor to the **use** of the data in a product or service or a particular business application (in particular, the exploration, identification, and validation stages of the analysis to determine the viability of the "research finding.")

Today, from a regulatory standpoint, the processes facilitating the knowledge creation phase or Corporate Research are generally treated the same as the processes facilitating the knowledge application phase. However, there is a far greater potential for impact (including risk) to an individual in the application phase than there is in the creation phase. By conflating the two phases, risk management is impacted disproportionally, i.e., higher risks associated with the application phase are applied to the lower risk phase of research. By extension, the value creation potential of data driven knowledge creation, even when there is a clear benefit to people, is suboptimized.

While the adoption and use of AI increasingly will require a separation of data driven development phases (AI has a much more pronounced experimentation stage than conventional analytical processes), most organizations currently do not differentiate clearly their data driven activities between the "identification" of the solution and the application of the solution. This problem is reinforced by current laws that do not support the split. These challenge organizational abilities to use data directly for Corporate Research, or that are less clear on allowing Corporate Research. This lack of clarity on the allowance of Corporate Research feeds the natural risk aversion for these organizations. For example, the term "scientific research" is narrowly defined and applied under the GDPR, making it less useful as a means for legitimate data processing in Europe, and is defined inadequately or has limited application as a data processing activity in proposed legislation elsewhere. This narrow application of Legitimate Interest, overuse of Consent and/or lack of clarity in proposed new laws relative to data driven Corporate Research and/or aggressive deidentification of personal data requirements feeds the natural risk aversion of some organizations to more aggressive data use.

The IAF's analysis suggests clarifying and enabling legitimate data driven Corporate Research through a combination of public policy adjustments and defined constraints or specific requirements on organizations would enable more responsible data driven innovation. This combination would serve people (citizens) and would better support and enable many economies' digital strategies while appropriately mitigating risks to individuals.

Before addressing a solution suited to this problem, it is useful to delve deeper into each of the reinforcing parts of this problem.

Organizations:

Interviews with multiple organizations (see Appendix I) and several academics suggest there are several approaches relative to organizations' uses of data in analytical processes. Since organizations are not monoliths, it makes sense to consider the most common of these approaches.

There is a continuum based on risk aversion (the EU is at one end of the spectrum, the U.S. is at the other end, and Canada is somewhere in the middle):

I. Where laws or regulatory interpretations of laws clearly prohibit or narrowly allow use of data, organizations are risk averse.
II. Where laws or regulatory interpretations are less clear and therefore more ambiguous, organizations have institutionalized prohibitions on data use based on perceptions of what court/regulatory interpretations will be, and this self-constraint adds some level of self-created risk aversion.
III. Where laws or regulations do not restrict to a significant extent the use of data, organizations are willing to assume risk.
IV. Where contracts contain restrictions on data use, based on (I) and/or (II), there is some additional level of risk aversion.

Organizational approaches to analytical processes differ but, in most cases, are not delineated clearly by lifecycle stages. Some approaches have a notional distinction between knowledge creation and knowledge application[4], but for most, data is used to solve a business problem. So, there is a presumption that data will be used even during the notional knowledge creation phase.

There are additional jurisdictional/organizational hurdles to the use of data. While in some cases there is no added utility of using identifiable data, in some scenarios the efficacy of the outcome simply is better if identifiable data can be used in the research phase. There are clear examples/data use scenarios where data has a people beneficial component and really requires identifiable personal data and de-identified or anonymized data will not generate the best result or even a useful result. As an example, while de-identified data is useful to recognize anomalous patterns and behaviors, in order to get more specific results, identifiable personal data is needed in the research stage.

[4] Some of these organizations have started to develop additional controls that are aligned with the Proposed Solutions detailed later in this paper.

A very common example of an anomalous fraud pattern is the credit card of a person who resides in country X being used to make an in-person purchase in country Y. Most likely that is a fraudulent transaction, unless the credit card company knows the cardholder is traveling in country Y.

However, identifiable personal data will be needed in the research stage, for example: in order to build an algorithm that looks for credit cardholders who no longer are customers (just looking for the 16-digit credit card pattern will not work because both current and former credit cardholders have the same 16-digit pattern, looking at expiration dates will not work because cardholders cancel their cards before the expiration date. So, in order to tell whether the algorithm works, real data or dummy data will be needed).

Another example involves the emerging "Metaverse." In order to build an identity management strategy for the metaverse based on, for example, voice characteristics, identifiable personal data will be needed. In the physical world, individuals' identities are validated in person (e.g., showing driver's license to open checking account at bank); in the digital world, the same information is provided over the internet as has been provided in the physical world and it is matched up; in the metaverse, individuals' identities are avatars and so personal information is needed to tell whether the avatar is the person the organization thinks it is and therefore that its identity management strategy works. In order to tell whether the algorithm that powers the identity management strategy for the metaverse built on voice characteristics works, the organization needs to have voice samples of real people and to be able to match those voice samples to real people.

There also is some evidence that the growth and maturity of an organization's risk-based governance has added to its general conservative approach; they have educated their businesses to the point where the businesses are less likely to take on risk. The result is less innovation based on an institutionalized approach to risk aversion. A practical implication of this risk aversion is that organizations face many data architecture challenges, and the investment required to address these challenges is increased or stymied by the public policy constraints (real and emerging).

Legal and Regulatory

There currently are laws that do not support the split (Corporate research and knowledge application) and challenge organizations' abilities to use data directly for Corporate Research or are less clear on allowing Corporate Research. This lack of clarity feeds organizations' natural risk aversion.

For example, while "Scientific Research" under the GDPR should encompass both public and private entities[5] and while the Article 29 Working Party Opinion on the Notion of Legitimate Interests states that "processing for research purposes (including marketing research)" could constitute a legitimate interest as long as the controller implemented sufficient safeguards, these interpretations have not been recognized by regulators. The EDPB has narrowed further what would be "scientific," by stating that the concept should not be stretched beyond its common understanding and that in particular, scientific research should refer to projects run "in accordance with relevant sector-related methodological and

[5] According to Recital 159, processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research. See also Articles 5(1)(b) and 6(4), Articles 9(j) and 89(1) of the GDPR, and Recitals 52 and 53

ethical standards, in conformity with good practice."[6] A recent Ropes and Gray blog pointed out, "Applications and interpretations of GDPR fail to consider adequately how research uses of personal data differ from other types, particularly as the data are pseudonymized."[7] In addition, the weaknesses in the utility of "Legitimate Interests" to justify data processing has led to further reliance on consent as a legal basis for data processing.

Under Canada's current privacy law, Personal Information Protection and Electronic Documents Act (PIPEDA), organizations can process personal information for legitimate purposes based on implied consent as a condition of the supply of a product or service. However, this limited processing may not provide the degree of clarity or scope of potential data use necessary for Corporate Research.

U.S. state privacy laws allow the use of personal information for research purposes only for limited purposes. The California Privacy Rights Act (CPRA) allows the use of personal information for technological development and demonstration, and otherwise must be consistent with what an average consumer would expect when the personal information was collected. Connecticut and Utah allow the collection, use, or retention of personal data to conduct internal research to develop, improve, or repair products, services, or technology.[8]

Future laws are not as restrictive with respect to research. While lack of clarity can create opportunity, it also can create uncertainty in some instances. For example:

- Canada's Draft new privacy law,[9] C-27 would be the first legislation to include both privacy and AI in the same overall law (different statutes), which may facilitate a more seamless recognition that AI requires substantial data and, by extension, that laws need to align on the treatment of data. This proposed law allows an organization, without an individual's knowledge or consent, to:
    - **Collect or use** personal information if the collection or use is made for the purpose of an activity in which the organization has a legitimate interest that the organization has previously determined outweighs any potential adverse effect on the individual from that collection or use. As observed in an article by the Canadian law firm Borden Ladner Gervais (BLG), it is unclear what specific types of activities fall within an organization's legitimate interests and whether an organization's legitimate interests include product improvement or development of new products or services. As with PIPEDA, this determination relies on "what a reasonable person would expect."
    - **Use de-identified information** for internal research, analysis and development, and select socially beneficial purposes. "De-identified" means to modify personal information so that an individual cannot be directly identified from it, even though a risk of the individual being identified remains. Personal information that has been de-identified for these purposes is not considered to be personal information and is therefore outside the restrictions in C-27. However, there is some ability to use identifiable data where it is clear the efficacy of the outcome would be better if identifiable information were used and subject to the organization taking into account

[6] EDPB Guidelines on Consent under Regulation 2016/679 ¶ 153 Version 1.1 Adopted on 4 May 2020
[7] GDPR and clinical research: time for a rethink? Edward Machin (ropesgray.com)
[8] The Connecticut Data Privacy Act and the Utah Consumer Privacy Act
[9] Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act - C-27 (44-1) - LEGISinfo - Parliament of Canada

the risk of harm to the individual that could result from using or disclosing the information.[10]

- U.S. draft American Data Privacy and Protection Act (ADPPA)[11] allows a covered entity to collect, process or transfer covered data to conduct internal research or analytics to *improve products and services,* leaving unclear whether the development of new products or services is included.

In addition, regulators need to become more comfortable with organizations using and sharing data, even de-identified data, for research purposes, particularly for people beneficial research purposes. For example, in order to better understand the impact of climate change, it is useful to have not only weather information but also insurance information, often personal information, which shows where claims have been made due to the impact of weather-related events. In this scenario, multiple data providers would need to participate. The law and regulators need to allow the appropriate Safeguards so Corporate Research for these kinds of socially beneficial purposes can take place.

Limiting the purpose of the research or the ability to use personal information in research does inhibit Corporate Research and/or does create uncertainty for organizations as to how or if to proceed. This limitation may have a negative or stifling impact on the creation of knowledge through data that may be very beneficial to people. While today, many organizations can accomplish their goals by using some version of non-identifiable data, it is likely future scenarios increasingly will be dependent on the use of identifiable data, particularly in research. It also is likely, as in the impact of climate change scenario, there increasingly will be examples of "pooled" data from multiple organizations that will enable research. Public policy needs to be forward-looking, or it always will be out-of-date.

The current state of regulatory limitations, organization risk aversion, and emerging public policy gaps reinforce the conflation of knowledge creation data activities with the application of this knowledge despite the different impacts and risks to individuals. This result in turn is sub-optimizing the benefits that could accrue not just to individuals but to society as a whole. The IAF thinks that organizations might be able to use personal data more broadly for Corporate Research if, in addition to legal and regulatory modifications, appropriate Safeguards were put into place and proposes that those Safeguards be developed and implemented in a defined "Research Sandbox" environment.

Potential Solutions

Public Policy/Regulatory

- **Expand Legitimate Interests** - In the EU, application of further processing under Article 6(4) and Legitimate Interests under Article 6(1) may be a solution with respect to Corporate Research if sufficient Safeguards have been implemented. Even if the Legitimate Interest legal basis could be used, it does not apply to the use of sensitive data (however, Recitals 52 and 53 suggest sensitive data could be processed where it is in the public interest to do so). In Canada, C-27 clarifies that Legitimate Interest expressly allows the use of personal data for research.

[10] Section 22(2) Government Bill (House of Commons) C-27 (44-1) - First Reading - Digital Charter Implementation Act, 2022 - Parliament of Canada

[11] H.R.8152 - 117th Congress (2021-2022): American Data Privacy and Protection Act | Congress.gov | Library of Congress

- **Expand "Public Interests"** – Assuming there is a boundary around knowledge creation vs knowledge application, where certain Corporate Research activities provide a clear public benefit
- **Clarify the ADPPA** - to include the use of personal data with respect to Corporate Research if it would improve the results significantly (it currently is limited to improving products and services)
  - **Leverage appetite for "Codes of Practice,"** for example as part of Bill C-27 in Canada - to further facilitate the use of all data for Corporate Research and organizational confidence.

The potential solution involves defining "sufficient Safeguards" (see below)

- In the EU, the legal basis for processing for scientific research purposes can be legitimate interests under Article 6(1)(f) or further processing under Article 6(4). If sensitive data is being used, then sufficient Safeguards need to be implemented.
- In Canada, C-27 might be revised so that research (including sensitive data) can be used without de-identification if sufficient Safeguards, including a Code of Practice, are put in place.
- In the U.S., the draft ADPPA might be revised so that research can be done for a socially beneficial purpose if sufficient Safeguards are put in place.

Organization – Practices and Accountability

There are 4 related conditionals that will be core to support public policy adjustments and could be built using the metaphor of a "Research Sandbox Environment".

1. **A defined beneficial purpose** and a defined way to put parameters around "people beneficial" and/or "Demonstration of Good Intent" must be determined as part of a required assessment.
2. **An organizational "code of practice"** must be created and adopted
3. **A defined and demonstrable accountability/governance process** addressing the Corporate Research stage of data use must be implemented. It is suggested that this process be modeled after Ontario's ICES[12] and Ontario's Personal Health Information Privacy Act (PHIPA)[13] requirement for the organization to establish policies and processes. This process must include a specific assessment that evaluates the full range of interests and risks of all stakeholders and mitigates risks as much as possible.
4. **The use cases and accountability processes the organization has committed to relative to data driven research** must be subject to some form of transparency.

Background Information Regarding Conditionals

[12] ICES is Ontario's independent, non-profit research organization that uses population-based health and social data to produce knowledge on a broad range of health care issues. Because ICES is a designated prescribed entity under Ontario's Personal Health Information Privacy Act (PHIPA), physicians and hospitals can disclose personally identifiable information to ICES with obtaining individual consent

[13] PHIPA requires the organization to develop policies and processes that are shared and reviewed with the Regulator. MANUAL FOR THE REVIEW AND APPROVAL OF PRESCRIBED PERSONS AND PRESCRIBED ENTITIES (ipc.on.ca)

- To differentiate knowledge creation from knowledge application, an organization must be able to demonstrate conditionals for their own "research sandbox" environment
- **Purpose**
    - Corporate Research cannot be open ended; Corporate Research must be "purpose driven." Defining the limits of what cannot be done with the data or insights (absent some other action) will be challenging. The base line assessment must be clear on the purpose (e.g., "to improve access to health-related services" or "to improve access to credit")
    - The purpose must be clearly aligned to a benefit to people, or corporate profit will be seen as the driving motive
    - The Corporate Research purpose must fit a public benefit (see UK ICO Sandbox criteria).[14] This requirement is to help compensate for pushback that an organization's expanded ability to use data is too broad

        ICO Criteria: ***A product or service is likely to bring some benefit to the public beyond the benefit of the submitting organization. A reasonable attempt is made to quantify potential public benefit in terms of breadth and depth and supported with some quantitative and/or position around "Demonstration of Good Intent"***
- **Governance**
    - An accountable person for oversight is necessary
    - Policies and procedures defining the requirements the organization has put in place regarding Corporate Research and what is required before any application of the knowledge (sandbox with a high fence) must be documented
    - A defined and demonstrable accountability/governance process must be developed and documented and available to a Regulator upon request
    - **A multi-stakeholder proportionality assessment must be conducted that takes into account the benefits, risks and interests of all stakeholders and appropriately mitigates those risks**
    - If de-identified data would sub-optimize the output, then identifiable data can be used as long as an analysis demonstrates the sub-optimization. **If personally identifiable data is to be used, then the assessment must demonstrate why personally identifiable data is needed and must include what risk mitigations have been put into place**
    - If, for example, sensitive data is being used or the intended outcome is in a sensitive class, the assessment must clearly identify the benefits and additional controls implemented (for example, addressing potential bias).
    - If de-identified data is used, a reasonable and defined state of de-identification (this can/should include policy and/or technical/process means to keep the data de-identified) must be met
- **Code of Practice**
    - An agreement to meet the requirements laid out in a Code of Practice (to be developed) could be reached
    - If a Code of Practice is being pursued,
        - A separate set of controls/ commitments (e.g., GDPR Article 40) must be developed and adhered to start to address the "trust issue."
        - Standards regarding processing controls (e.g., BCRs) must be created and adhered to

[14] Sandbox assessment Criteria Indicators (ico.org.uk)

- **Transparency**
    - There must be publicization of process (risk management) and demonstration of types of data uses and decision making (demonstrable accountability) in Corporate Research.
    - There must be additional transparency (e.g., public facing website) as to research data practices.
    - There must be some form of public transparency regarding the use cases and the processes to which the organization has committed.

Conclusion

The successful advancement of digital agenda will require the creation of safe pathways for data to be turned into information, for information to be turned into knowledge and for data to facilitate actions that are societally beneficial to people, all where processing is conducted in a manner that is lawful, fair, and just.

In order for these pathways to develop, the current paradigm will need to evolve in at least three ways:

- Recognition that data driven knowledge creation or "research" increasingly will be generated through advancement's made by corporations and not just academic institutions. As data becomes an ever-increasing asset for value creation by organizations, the use of advanced data intensive technologies, such as AI and Machine learning, will be used by organizations that have multiple objectives. This recognition that knowledge creation and knowledge use can have value for both (i) businesses and shareholders, and (ii) individuals and groups of individuals will remove unnecessary impediments to the digital revolution. The IAF has a sister project on multi-dimensional proportionality that is exploring how the full range of interests and stakeholders can be more thoughtfully balanced with individuals' privacy interests.
- Placement of requirements by public policy and by extension regulation on data processers that are gated on the application of data driven knowledge. Today the digital age is accelerating, with some organizations using data without more broadly assessing the adverse consequences associated with that use. This failure has led to a stronger hardline on data use by policymakers and regulators.[15]. Today, much of this focus has resulted in creating restrictions that have had the by-product of restricting the use of data for research, even where there is limited risk to individuals. Factoring in the people beneficial knowledge that comes from knowledge creation would be useful.
- Evolution of accountability in two related ways:
    - If data is to be better used for the creation of knowledge, then there needs to be additional constraints that are supported by increased accountability by organizations wishing to avail themselves of this potential. Further, these constraints need to be specific to the phase of data driven knowledge creation and separate from the requirements should the knowledge be put into action. Organizations will need to evolve their capacity and capability to more effectively assess the risks (and benefits) to multiple stakeholders and not just to themselves. These capabilities will need to be demonstrable and demonstrated.

[15] Republicans and Democrats, Unite Against Big Tech Abuses - WSJ

- Regulations need to evolve to measure and enforce demonstrable accountability that enables data to be used for maximum knowledge creating benefit and that are commensurate with the risks to individuals and groups of individuals at this stage of data processing.

The IAF appreciates that this level and type of paradigm shift will require input from key stakeholders, including regulators and organizations. The IAF hopes to obtain this input through this project which includes workshops in 2023.

Appendix I

The following organizations provided valuable insights to this project:

Acxiom

ATT

Ceridian

CISCO

Citrix

Dennis Hirsch – Ohio State University

Facebook

Inst for Clinical Evaluation Studies (ICES)

IPG Brands

Loblaws

Mastercard

Merck & Co., Inc., Rahway, NJ, USA

Michael Geist – University of Ottawa

Sun Life

TELUS

Appendix II

Information Policy Pulled by Two Extreme Poles – Creating Tangible Risk for All Stakeholders

Polarization seems to affect the public policy debate on the world of privacy and data protection. Terms such as autonomy and control increasingly are being interpreted as personal sovereignty. Policymakers designing the new data-driven industrial policy are talking past independent regulators who are concerned markets are dominated by data extraction that harms all individuals, particularly protected classes. This polarization is making legislation and regulation that both facilitates innovation and protects the full range of stakeholder interests increasingly difficult. Advanced analytics, including AI, require use of data pertaining to people. Emerging laws, rules, and regulatory interpretations driven by polarized views increasingly make using data difficult. The chart below describes this polarization.

"People are at Risk due to Commercial Observation, Analytics and Data Monetization" (*one perspective*)	**"Countries and Economies are at a Competitive Disadvantage if Data is not Used to Fuel Growth and Solve Problems" (*another perspective*)**
Pole One - Individual Sovereignty, a Predominant Fundamental Right in the Observational Age, is at Risk from Data Extraction • Mobile and online tracking is pervasive, and IoT technologies are accelerating a hyper-observational trend, leaving people with no place to be free of being watched (and being monetized) • Advanced analytics and AI are accelerating the power asymmetry between people and institutions, for both small and large organizations, and therefore are heightening the geopolitical balance of power perceptions • Technology links people to their many connected devices and collects the data generated from them – making the data usable in ways that don't always serve the data provider. • Advertising and other tracking data uses have been personalized to the point that they are experienced as manipulative rather than persuasive	**Pole Two – Knowledge Creation and Application are the Drivers of Digital Innovation and Progress, and Market Forces will Correct Harms** • From a national perspective, digital technologies will drive the competitive position of national economies and help grow the total global economic pie. This approach will deliver better economic opportunities for people as well as improved educational and healthcare outcomes. • From a corporate perspective, data, analytics, and AI will drive competitive advantage, serving all corporate stakeholders. This result, in turn, will deliver better products, improved access to services, better health outcomes as well as broader societal benefits. All will be enhanced by the application of AI. • Data-driven advantages and benefits are dependent on data being available for knowledge creation. • The countries that harness the digital future will require data-driven knowledge creation be conducted in a responsible

Policy Trends that Support Greater Individual Sovereignty

- Biden has ordered the FTC to establish a rule on ***surveillance*** and data ***accumulation.*** Dr. Shoshana Zuboff, author and Harvard professor emeritus has coined the term surveillance capitalism. Speeches by FTC commissioner Slaughter indicate that work on surveillance capitalism is the foundation for "the new FTC" privacy rulemaking.
- Dr. Zuboff received the Buttarelli Award at the 2021 Global Privacy Assembly for her work describing "surveillance capitalism," indicating the power the concept has with global regulators.
- The European Data Protection Board has indicated that all paths to AI and data innovation go through the GDPR.
- The Ontario Government proposes to regulate private sector privacy in line with the province's AI ambitions. The Ontario Information Commissioner has suggested that research be limited to where *the research purpose is intended to advance the public interest.*
- The undefined terms "*Fair" and "Unfair"* are in almost all privacy legislation and by extension become enforceable. The usage of this terminology is translating into procedural restrictions where regulators subjectively sense data use is unfair.
- Data science is seen as too complex to ever be transparent enough for individuals to be able to exercise the right to object and therefore be lawful, leaving no legal means to do data science.
- AI regulation is seen as in support of economies' digital agendas and in conflict with the application of data protection regulations. *This tension creates competing conflicts for organizations and governments.*
- AI ecosystem audit requirements will evolve but will remain an unsatisfactory and trustworthy fashion, raising the questions of what constitutes responsible and trustworthy.

Policy Trends that Support the View that Digital Innovation is key for Global Competitive Advantage

- U.S., EU and other national/regional digital policy agenda
- Chinese policies supporting AI competitive advantage
- UK DCMS Consultation on updating UK GDPR
- Public sector AI solution procurement requirements in the EU AI Regulation will have impact beyond EU borders and have the potential to create an AI audit ecosystem (see Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation | SpringerLink)
- Singapore regulatory reforms to encourage greater innovation with data pertaining to people
- Think-tank references to the Fourth Industrial Revolution
- References to the "data imperative" by major consulting firms

solution to the ethical application of AI and associated data sources.	