



Research Data Use in a Digital Society: **A Deliberative Public Engagement**

October 19-20 and November 2-3, 2019
Vancouver, BC
University of British Columbia

This booklet was prepared by the BC Data Deliberation research team:

*Jack Teng
Michael Burgess
Kimberlyn McGrail*

*Colene Bentley
Kieran O'Doherty*

September, 2019



IMPROVE LIFE.



Table of contents

1	What is the purpose of this booklet?	4
2	Building on the recommendations from a previous public deliberation	5
3	What is a public deliberation?	5
4	What is the importance of this public deliberation?.....	6
5	What happens during a public deliberation?.....	7
6	The increasing availability and collection of data.....	8
7	What types of data are currently collected?	9
8	New data sources and challenges	11
9	How or what governs the collection and use of these data?	13
10	What authorizations are needed to share data?	13
11	What are some examples of research with potential public good?	14
12	What are the potential risks?	16
13	Managing future uses and applications of data research	18
14	Summary	19
15	Your role in the deliberation	20
	Glossary of Terms.....	21

1 | What is the purpose of this booklet?

The information in this booklet comes from academic literature, consultations with experts and stakeholders, and media such as newspapers and radio.

The intent of this booklet is to inform you about the collection of **individual-level data** in our increasingly digital society. It will explain how technology is creating opportunities for **research** by generating new data and more groups interested in using data. It will outline both potential benefits and potential risks of these developments. It also describes current legislation and policies that guide how data can be shared for research.

The information in this booklet is intended to support you and your fellow participants in the public deliberation. We hope it will also be useful for ongoing discussion and reflection on these and related topics. You are not expected to be experts on this topic, and you will likely have further questions after reading this booklet. We encourage you to bring these questions to the deliberation, along with your insights and perspectives. These are all critical to the deliberation.

A glossary can be found at the end of this booklet. Glossary terms appear in bold lettering throughout the booklet.

Please feel free to share this booklet.

| *We encourage you to bring questions to the deliberation, as well as any insights and perspectives you may have.*



2 | Building on the recommendations from a previous public deliberation

The public deliberation you are participating in builds on a previous deliberation held in Vancouver on April 7-8 and 21-22, 2018. That event, titled *Using Data About You for Research: Who, How, and Why*, examined existing data research practices and regulations.

During this event, you will be considering new potential applications, uses, and partnerships in data-based research and the rules and regulations needed to ensure their trustworthiness. Some of these new applications, uses, and partnerships have not yet been tested or implemented. In some cases, legislation and policies are in development to regulate these new possibilities. The results of this deliberation, and thus your contributions, will be used to inform these policies.

Participants in the previous deliberation developed recommendations aimed at ensuring the trustworthiness of existing procedures, regulations, and governance of how **linked data** are shared for research purposes. Generally, participants were supportive of research using linked data because of the potential societal value of such research. Participants expressed a desire to see the data access request process made more efficient to facilitate more research, as long as there are adequate protections in place. More information about this event, including a summary report that includes the recommendations, can be found here: <https://www.popdata.bc.ca/BCDataDeliberation>

In preparation for the event, and over four days of the event, you will be given written materials to review, hear expert and stakeholder speakers, explore your perspectives and that of other participants. You will then consider how to balance the diverse interests related to research data use (and specifically **linked data**) in a digital society.

3 | What is a public deliberation?

A **public deliberation** is a community discussion on issues that affect members of the public such as yourself. People from British Columbia are selected to reflect the diversity of life experiences and perspectives in our province. A public deliberation is a democratic process that supports citizens to understand issues and different perspectives about those issues. The desired outcomes include recommendations and/or identification of areas of disagreement among participants. Policy makers, experts, and stakeholders may provide information or attend the deliberation as observers.



4 | What is the importance of this public deliberation?

Public deliberations are discussions about important societal issues that involve values or trade-offs. Instead of *telling* the public how such issues will be resolved, deliberations invite the public into *active participation about the issue*. Members of the public have an opportunity to identify what is important to them about a societal issue and provide advice, in the form of recommendations, to policy makers. In this deliberation, the trade-offs involve balancing the potential benefits of research using **linked data** with potential risks, such as privacy.

| *A public deliberation is about respecting the diversity of perspectives amongst us and finding ways we can live together.*



A public deliberation is about respecting the diversity of perspectives amongst us and finding ways we can live together. The information you read and hear may inform your opinions, and your opinion might (or might not) change over time. The intent is to inform and engage citizens as they discuss issues and make recommendations.

5 | What happens during a public deliberation?

On the morning of the first day, you will hear more information from experts on different aspects of using data for research. The speakers are chosen to provide a wide range of perspectives and will not necessarily agree with each other. You will have the opportunity to ask them questions to clarify what they said and get more information.

For the remainder of the four days, you and the other participants will work with trained facilitators in both small groups of 6-8 participants and the full group of participants. You will be discussing a series of questions that the organizers have prepared for the event. For each question, you will first discuss the different perspectives and the reasoning behind your perspective in your small groups. You will then return to the large group to consider the different viewpoints raised in the small groups. The small groups are meant to bring out the broadest range of perspectives. The large group is intended to understand that diversity and work together to imagine how it can be accommodated in our province.

In the large group, the facilitator will follow the participants' lead on any emerging consensus of opinions. Based on this, the group will formulate policy recommendations. While the recommendations are important because they represent agreement, there may be times when you and your fellow participants do not all agree. In all cases, we are interested in the reasons behind your perspectives, as this is also important information for policy-makers, particularly where there is no consensus.

On the last day, participants will have an opportunity to review all of the recommendations together and make any necessary modifications, and then to present those recommendations to a panel of policy-makers.

Example deliberation questions

What are the qualities of research partners that make them trustworthy?

What kind of approvals need to be in place for data research to proceed?



6 | The increasing availability and collection of data

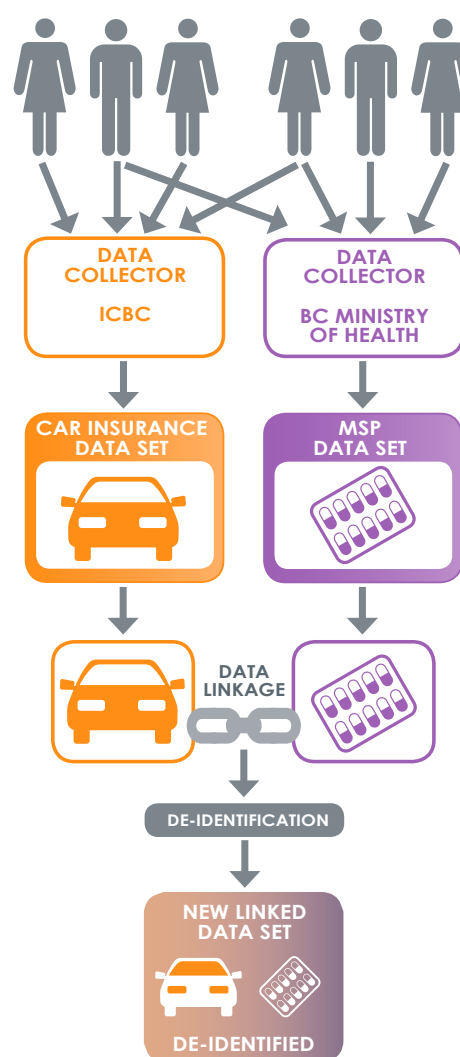
The expansion of computers and other information technologies like smart phones are increasing the amount of **data** that are collected about us. Many aspects of our lives are being recorded, tracked, and turned into data. Everything from our daily commute, to the time we wake up, our levels of exercise, and even the type of coffee we enjoy might be collected by or through various public and private entities, such as researchers, governmental agencies, doctors' **private medical practices**, apps, and devices like Fitbits. Since these new data are digital, they are more readily accessible, and can easily be moved, shared, and used, including for research purposes. In addition, the time between obtaining research results and implementing them is narrowing, which increases the possibility for research to impact us directly.

Let's take a look at an example.

Individual-level data are collected from a single source, such as the Insurance Corporation of British Columbia (ICBC), and together they constitute a **data set**. Data sets from different sources (e.g., ICBC and Medical Service Plan (MSP)) can be shared and combined to create **linked data**. ICBC collects data from individual drivers in BC each time a person renews their car insurance. MSP collects data on visits to physicians. These data can be combined to provide insights on the relationship between health care services use and driving.

Linked data sets are of particular interest to researchers, since the broad range of information they contain can shed light on complex issues. When linked data are used for research purposes, they are **de-identified** to reduce the possibility that they could be traced back to any specific individual. For example, linking data about early childhood, education, and employment could help researchers understand the long-term effects of early childhood socio-economic conditions, and where to intervene to improve longer-term outcomes.

New sources of data mean that new types of linked data sets can be created. The new linked data sets offer new opportunities



not only for research, but also new partnerships between different entities, new uses of data sets, and new applications of the results derived from the data. These new opportunities can lead to insights that benefit society. However, they can also lead to new risks to individuals and/or communities, or could be used to make controversial policy decisions that impact communities as a whole or place vulnerable populations at risk. This deliberation will help develop recommendations that balance the risks and benefits of research using linked data.

7 | What types of data are currently collected?

In British Columbia, data are collected by both individuals and organizations, such as **public bodies**, **researchers** and **private enterprises**.



Public bodies comprise a wide range of federal (e.g., Canada Revenue Agency, Statistics Canada) and provincial organizations (e.g., ICBC, Ministry of Health, Health Authorities, Ministry of Education, Universities). Public bodies collect personal information such as demographics, driving records, workplace injuries, educational attainment, and income levels.



Researchers are people who conduct statistical and other analyses of data. Researchers often work at universities, Health Authorities and other **public bodies**. Researchers collect a wide-range of data for their investigations, including both clinical and survey data.



Private enterprises include corporations like Amazon, Loblaws, Facebook, and Google, Apple, and app developers. Private enterprises also include pharmaceutical companies, device manufacturers, and many doctors' offices. To use their services, we often provide private enterprises with our personal information, such as credit card information, phone numbers, addresses, and date of birth, along with any data about service use or purchases.

The data collected can be categorized into the following types:



Administrative data: data collected in the course of providing and/or paying for services (e.g. hospital admissions, physician payment information, immigration records). This type of data can also be collected when you shop, when you drive a car, or when you use an online app (e.g., Modo car sharing; Shaw; Netflix).



Clinical data: data that includes specific aspects of persons, conditions, and/or care (e.g. blood pressure, lab results, diagnoses, clinical notes, genomic information)



Survey data and data from research participation: data collected directly from and about individuals or groups, which may be used for research (e.g., age, ethnicity, education, income, daily activities, opinions)



Licensure/registry data: data about specific groups of people for regulatory or monitoring purposes (e.g., Vital Statistics, professional regulatory bodies such as the College of Physicians and Surgeons, cancer registries)



Physical and biological data: data about specific aspects of human biology (e.g., height, weight, blood pressure, blood or tissue samples). When these samples are collected in a central location, they are sometimes called “biobanks.”



Surveillance data: data about individuals in the context of everyday life that can be used to track locations, behaviours, and activities (e.g., CCTV video capture, Fitbit, web activity, GPS tracking, grocery purchases, car sharing)



Internet of things: data that are generated by “smart” appliances or other technology that may be found in cities, private residences, and institutions. This is an emerging data type that is expected to grow exponentially in the coming years.

Data can fit within more than one category. For example, physical and biological data could be collected during a research study, and would thus also be considered survey data and data from research participation. Similarly, administrative data collected about purchases online and coded based on location could also be considered surveillance data.

8 | New data sources and challenges

Data resulting from digitization

New sources of data are being collected by both public bodies and private enterprises. The data they collect are not restricted to a specific type of data, but can encompass any or all of the types of data listed above. The difference is that as our society becomes more digital, more data are being collected by public bodies and private enterprises with greater ease, and sometimes without our knowledge. The uses of these data are often not all anticipated at the time of collection. In fact, uses of data are constantly evolving along with new technology and research techniques. Using data for a purpose other than the original intent is known as a **secondary use of data**.

Health-related devices increasingly include the ability to collect data as they are being used and then send those data to manufacturers. Diabetes devices, such as blood glucose monitors, continuous glucose monitors (CGMs), insulin pumps, and other wearable sensors, are examples of this. To ensure that they are working properly, the devices will collect administrative data related to the number of times the patient uses the device (this would be called administrative data and surveillance data), the doses the patients may require (this would be clinical data), and biometric information, such as the patient's heart rate and blood pressure (physical and biological data). These data are then sent to the device manufacturer (i.e., a private enterprise) for purposes such as quality control or device improvement.

Research using data from such devices could be informative, especially if those data are combined with other data sets, such as health administrative data or clinical data from **electronic medical records**. This would be a secondary use of the data. Another example would be to combine data collected from apps or devices (e.g., Google Assistant; Amazon Echo; FitBits; health apps) and with other types of behavioural or health research data, such as age, other demographics or education. In both of these examples, combining these data would require that the data sets from private enterprises be linked with data sets from public bodies.

The role of legislation

The ability to combine different types of data is governed by legislation. The two main pieces of legislation in British Columbia are the *Freedom of Information and Protection of Privacy Act* (or FIPPA, for public sector data) and the *Personal Information Protection Act* (or PIPA, for private sector data). These two acts have many similarities, including allowances for disclosing personal information, without consent, for research purposes.

The proper application of PIPA and FIPPA is the responsibility of the Office of the Information and Privacy Commissioner, an independent government organization. Independent government organizations are non-political organizations that monitor governmental practices and ensure they are legal

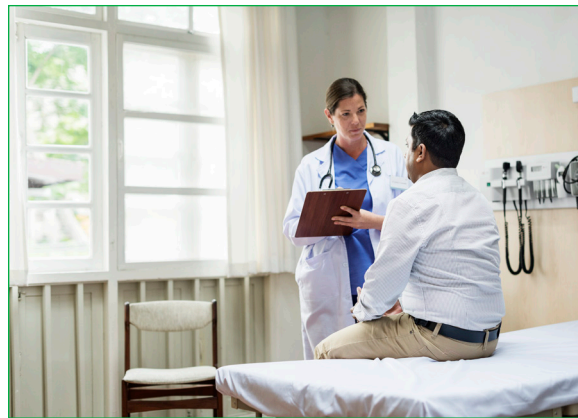


and ethical. Complaints can also be filed through that office if an organization is thought to have used data improperly. **There are currently no provisions within PIPA or FIPPA about how public sector and private sector data can be used together.**

Challenges with combining public and private data: the health care data example

The difference between FIPPA and PIPA can be seen in an example from health care. Physicians who work as private business owners in the community (most physicians in BC) fall under PIPA while hospitals and health authorities and the Ministry of Health fall under FIPPA. In practice, this means that electronic medical records held in physician offices are not automatically linked with data from other public bodies, such as hospitals or the Ministry of Health. Currently there is no centralized data set in BC with every patient's medical record, either from electronic medical records on their own or linked with administrative data.

| *System-wide access to multiple data sets might help individual patient care by enabling health professionals access to data from others about patients they are treating.*



There are potential benefits to harmonizing the regulations that govern health data from public bodies and private enterprises. Being able to access a broader variety of health data could facilitate research that informs quality improvement as well as equity and efficiency of the health care system.

The Ministry of Health is currently developing legislation that is intended to harmonize the rules that govern health data. How some of those rules will be written is still to be determined, including research uses of data. The recommendations from this deliberation will be used as input to that legislation.

9 | How or what governs the collection and use of these data?

When it comes to data collected by public entities, each data set is generally administered by an individual **data steward** whose role it is to ensure that the standards of FIPPA are upheld. Within these public bodies collecting data (e.g., health authorities, hospitals, provincial agencies, ministries) a data steward manages the data set and makes decisions on whether the data can be shared, with whom, and whether the data may be combined to form linked data sets. Data stewards follow procedural rules and policies determined by the public body. Those rules and policies will be consistent with FIPPA as well as any other legislation that might apply, and in some cases will layer on additional protections which can affect the speed at which approvals are obtained to share and link data sets.

Generally, for a data set to be released from one public body to another public body (including to researchers), there is a review process that includes a review of a research proposal, and an expectation of review by a **research ethics board**. Both data stewards and research ethics boards consider the implications of any data linkage involved in these requests, including data that might come from private enterprises.

10 | What authorizations are needed to share data?

PIPA and FIPPA provide privacy protections for personal information, and place limits on how and when information can be collected, used or disclosed. They also specify how and when **informed consent** needs to be obtained – and when it does not need to be obtained – when collecting data, as well as what level of data security is required.

Public bodies are allowed to collect personal information for a variety of purposes, including those that are necessary for providing services. For example, WorkSafeBC collects personal information about workplace accidents that are reported to them. Consent is not required for this collection, or for example if someone requires emergency medical treatment and is not able to provide consent. The **secondary use** of those data for research purposes also does not require consent. The data steward responsible for the data set has the discretion to determine whether the secondary use proposed by the researcher is appropriate, guided by specific parts of FIPPA as well as ethics guidelines that govern research.

Private enterprises generally must have the consent by the individual to collect **personal information** (though that consent can be implied) and must provide an option to opt out of that collection where the collection is not required in order to provide their product or service. For apps or online services, consent for any secondary uses may be as simple as checking a box to authorize data collection or may be implied. There is no requirement to read the disclosure agreement, and researchers have estimated that it would take 76 work days to read in full all the privacy policies we encounter in a year. Similar to FIPPA, private enterprises can disclose data to researchers without consent under certain conditions, including that there will be no harm to individuals and the research is in the public interest. (FIPPA and PIPA are pretty similar in most of these expectations.)

Researchers based at universities must have informed consent to collect data from study participants, for example if they are conducting a survey or doing a focus group. The consent forms are reviewed and approved by research ethics boards. Different forms of consent may be required for different types of research projects. For instance, a study participant may agree to have his/her data used only for that study, meaning that additional consent would be needed for any secondary use of the data. Alternatively, the study participant may agree to **global consent**, which would allow for the research data (including personal information) to be shared with other researchers. This arrangement is useful for research projects that collect long-term data that can be used by many different researchers to answer many different questions (e.g., Canadian Longitudinal Study on Aging).



Canadian Longitudinal Study on Aging

The Canadian Longitudinal Study on Aging is a large, national, long-term study of more than 50,000 individuals who were between the ages of 45 and 85 when recruited. These participants will be followed until 2033 or death. The aim of the study is to find ways to help us live long and live well, and understand why some people age remain healthy as they age while others do not. Data collected are managed and stored centrally in a secure and protected environment.

11 | What are some examples of research with potential public good?

Research using linked data sets can lead to important insights that benefit society. The examples below describe research that has been completed using linked data, and research that could be enabled by using new sources of data.

Research using public sector data

The following is an example of research that was only possible using linked public sector data.

Dr. Mieke Koehoorn at UBC was interested in looking at how many people affected by a type of cancer called mesothelioma receive workers compensation. This cancer is caused by exposure to asbestos, and in about 95% of cases the exposure happens in the workplace. People who have mesothelioma and were exposed in BC are eligible for compensation through WorkSafeBC. The complication is that the disease develops very slowly, with symptoms generally appearing 30 to 40 years after exposure. The questions was whether there is a high rate of uncompensated cases in BC.

To examine this, Dr. Koehoorn used linked data that included cancer incidence data from BC Cancer and injury claims data from WorkSafeBC. The approval process involved review by a research ethics board as well as by the two public bodies. Once approved, the data sets were linked and de-identified by Population Data BC.

Dr. Koehoorn found that less than 50% of mesothelioma patients receive compensation from WorkSafeBC. She also found that women and retired workers are least likely to seek compensation.

These results had direct policy implications. BC Cancer and WorkSafeBC worked together to increase the awareness of compensation benefits for people with mesothelioma. They also knew to target women and retired workers

What is Population Data BC?

Population Data BC (PopData) is a multi-university data and education resource. PopData staff facilitate interdisciplinary research on the determinants of human health, well-being and development. Services include securely housing and managing data, linking data, de-identifying data, coordinating requests for access, and operating a secure research facility for data analysis.

Potential research using new sources of data

The following example, a controversial waterfront development called Sidewalk Toronto, shows how data from private entities could support new research opportunities.

Sidewalk Labs is owned by Google's parent company Alphabet, and they plan to invest up to \$1.3 billion to develop a 4.8 hectare property on the waterfront in downtown Toronto. The plan is to develop this area as a residential-focused "smart city" that will implement high-tech solutions to urban design, building, mobility, and sustainability. The project promises to improve many aspects of the residents' lives, such as reducing congestion, increasing safety, and introducing personal rapid transit solutions (e.g., self-driving cars).

| *The Sidewalk Labs "smart city" development aims to improve the lives of the residents who live there by collecting and analyzing detailed, individual-level data about them.*



To do so, Sidewalk Labs plans to collect continuously a large amount of highly-detailed individual-level data about the people who live in or use the neighbourhood. These data may include the number of steps people take outside their homes or how long they use recreational facilities. There is clear research potential for these data, for example to address questions about urban planning, health, and sustainability. Although the residents will consent to have data collected about them, critics of the project say that this type of data collection is problematic in part because a private company will have a role in the relationship between citizens and city government.

To address these concerns, Sidewalk Labs has committed to creating an independent data trust with a mandate to govern the use of the data collected in the public interest. This would include creating clear responsible data use guidelines and ensuring data are not used for other purposes. Critics raise concerns about the potential to collect very detailed, individual-level behavioural data, and that both the capabilities of such data and the proposed governance are untested. The risks in a sense are completely unknown.

12 | What are the potential risks?

Bringing a variety of personal and potentially sensitive data together inherently carries risk. While there are a variety of potential risks, we highlight three that are raised routinely:



Loss of privacy through data breaches

Organizations as well as the general public are concerned about the risk of a data breach, that is, when data are improperly released (disclosed). This might happen either by accident or through a malicious action such as hacking or theft. The results could range from anxiety of the people affected to the selling of stolen data, and to identity theft. A prominent example of a data breach happened in BC in 2012, involving the BC Ministry of Health and the personal information of 38,000 British Columbians.

A recent example of an accidental data breach involving private entities involved the voice-controlled device, Google Assistant. The device can answer questions and perform commands, such as report what the weather will be, play music, and make online purchases. As part of its quality control process, the device sends voice recordings of the owners' commands to be transcribed by human operators. A report has shown that sometimes the voice recordings can reveal personal and sensitive information about the owner.

These data breaches have been happening around the world, and security consultants believe the frequency of these breaches will increase.

Intended or unintended implications for communities and groups

No matter what security regulations are in place, there is risk that data could be used to develop insights that are then put to use for undesired or unethical purposes. One example of this is the use of data to develop algorithms that target and influence individuals, such as targeted Facebook ads using misleading information to influence the outcome of elections or referendums such as Brexit.



The research itself may also be problematic, in part because the research can only be as good as the data that are used to produce it. For example, it is known that African Americans have the highest mortality from skin cancer in the US; this is in part because physicians are not

AI vs algorithms

The term AI or artificial intelligence covers a wide-range of ways to analyze data, though all tend to use large data sets and high-power computers. The computerized process uses algorithms, that is, rules that human researchers develop.

trained well to detect cancer on darker skin. New methods using artificial intelligence to develop diagnostic algorithms are better than physicians at detection, but if those models are trained with data largely from people with lighter skin, this new diagnostic method will not correct fully for the existing bias.

Algorithms in action

One example of an algorithm was developed in China to assess public comments on social media and to identify key phrases that people who are suicidal may use. After identifying people who were at risk, social workers were able to extend support to these people and contact them via social media. While there is potential public benefit from this project, it also raises questions around surveillance and privacy.

Unanticipated future uses of data

The time that it takes to make policy or program delivery changes based on research is shortening. In fact, one of the motivations for more use of secondary data is precisely this – to make more policy decisions based on evidence. These changes are likely to modify some people's access to a service. Often, research can make the use of services more cost-efficient, for example by increasing the speed of access, or ensuring that services are more attuned to community needs. In other cases, people may lose eligibility and thus access to certain benefits they previously had, for example if we find that employment programs are effective for some age groups but not others.

Making policy decisions based on research evidence is a broadly shared goal because it is a way to ensure that services are of high quality, are effective, and are delivered fairly. This does assume that the research evidence being used is of high quality and does not include inherent bias against certain kinds of communities or sub-populations. Given that the pace of change both in the type of data available and the ways they are being analyzed, we will likely be facing ongoing questions about what kind of uses are acceptable and which may raise concerns.



13 | Managing future uses and applications of data research

Anticipating the possible uses and applications of data-based research and the positive and negative impacts they may have is challenging. To help address these challenges, decision makers could incorporate insights from people like you to make decisions that are acceptable to the people of BC.

Individuals and communities can provide insight in a number of ways: through surveys, focus groups, town hall meetings, health advocacy groups, and so forth. Public deliberations like the one you are attending are also used to help decision makers understand what is important to people about a given issue.

Models of engagement

There are several models of public engagement that are used by different organizations around the world. Some might involve fewer than ten participants and have a dedicated focus on a particular topic or issue, while others have 30 or more participants that meet once or twice a year over a few days to delve deep into a policy issue, revisiting the issue as needed.

Different models suit different communities and policy needs, and budgets. Whatever model is used, there are many pros and cons. An advisory that meets regularly might be very comfortable with discussing and debating the issues. However, as a group, they may start to become “experts” on the issue, thus moving further away from the general public’s viewpoint. This can be problematic for decision makers, who want to make sure they hear from the public about the issue.

There are ways to help prevent the advisory’s perspectives from becoming expert perspectives. For instance, membership terms may be established for the advisory, such that a new member can be brought in once an existing member’s term is over. This may help maintain the ‘public’ part of the public advisory.



14 | Summary

New and more data are being collected about us every day. New sources of data can be combined to create new linked data sets in new contexts and with new partnerships. This will create opportunities for researchers to pursue, and lead to new discoveries that positively affect individuals and society. Further, because of new technologies, the time that it takes to apply research results in practice is decreasing.

| *New sources of data can be combined to create new linked data sets in new contexts and with new partnerships.*



The same things that make new data-based research opportunities so promising and powerful also raise issues around maintaining privacy and the ethical use of data and the information that are derived from them. It is thus important to develop policies and regulations that can balance the importance of benefiting society and the risks of conducting and applying the research.

15 | Your role in the deliberation

During the deliberation, you will hear more about new data practices, sharing data, and privacy from speakers who have expertise on particular issues and the other deliberants.

You and your fellow deliberants will bring your own perspectives to the discussion. You are not expected to be an expert on this topic.

You will be asked to discuss some of the issues related to sharing and researching linked data with the other deliberants. These may include issues such as:

- Under what conditions is it acceptable to combine public and private data for research?
- What kind of authorizations need to be in place to share sensitive data sets?

We hope that you will bring your opinions, values, and ideas about data and privacy to the deliberation. You will work together to make recommendations that can be used to more effectively inform policy decisions on data access regulations.

To facilitate discussion, we ask that you follow these ground rules:



Keep an open mind



No eye rolling



Listen to others



Avoid cross-talk



Participate in respectful deliberation



Try not to interrupt



Ask for clarification



Try to justify your opinions

Glossary of Terms

Administrative data: data collected in the course of providing and/or paying for services (e.g. hospital admissions, physician payment information)

Algorithms: are a step-by-step approach to solving a problem. They can be written as a set of instructions for a computer, for example to use data to make predictions. A model or algorithm is often the outcome of a research project.

Biobank: a dedicated institution preserving biological samples (e.g., spit, blood, tissue). These samples are usually used for research, such as at the BC Children's and Women's Hospital.

Biometric data: data regarding the physical descriptors of an individual (e.g., fingerprints, retina scans)

Clinical data: more detailed information about specific aspects of persons, conditions and/or care (e.g. blood pressure, weight, lab results)

Data breach: an unauthorized release of data usually through hacking or by accident

Data set: a collection of data that has been gathered using the same criteria

Data steward: a designated official responsible with approving or denying data access requests

De-identified data: data where the personal identifiers of the individuals have been removed with the intent of minimizing the chance of re-identification

Electronic medical records: records in digital form kept by physicians or hospitals or other care providers. These records contain clinical data.

Genomic data: DNA-derived data about molecular aspects of human biology

Global consent: A type of consent where the individual agrees that their data be used for secondary uses without needing to ask for re-consent, but usually with constraints, such as the secondary use being for research purposes.

Individual-level data: data that are collected that can be associated to an individual

Informed consent: an explicit agreement by the individual to have their data collected under the terms that are described in a document.

Licensure / registry data: Information about specified groups for regulatory or monitoring purposes (e.g. Vital Statistics, professional regulatory bodies such as College of Physicians, cancer registries)

Linked data: a collection of data, usually for research purposes, that combines two or more different sources

Peer review: subjecting research plans to review and assessment by experts. This often determines whether funding is provided for the research

Personal information: is information recorded about an identifiable individual. It can include: your race or colour; national or ethnic origin; religion; age; marital status; blood type; fingerprints; medical, criminal or employment history; information on financial transactions; home address; and your Social Insurance Number (SIN), driver's license or any other identifying number assigned to you.

Privacy: an individual's right to be free from intrusion or interference by others. Privacy risks are the potential harms that participants, or the groups to which they belong, may experience from the collection, use, and disclosure of personal information for research purposes. Privacy is distinct from confidentiality which is an ethical and/or legal responsibility of individuals or organizations to safeguard information entrusted to them, from unauthorized access, use, disclosure, modification, loss or theft.

Private enterprises: any privately owned business or entity, such as a corporation, private clinical practice, union, or club

Private medical practices: physicians who work as private business owners in the community (which is most physicians in BC) are considered to have private medical practices

Public bodies: provincial or federal organizations, such as Ministries, agencies, universities, and hospitals

Public deliberation: a community discussion on issues that affect members of the public. It is a democratic process in which citizens participate, and policy makers and experts provide information and observe. The results of a public deliberation can inform important policy decisions.

Qualitative data: data that cannot be meaningfully counted, such as a category (e.g., favourite colour, strength of preference for a type of coffee) or text (e.g., responses to open ended survey questions)

Quantitative data: data that can be meaningfully measured in numbers, such as height and weight

Research: the development of questions and systematic investigation into them

Research ethics board: panels based at universities and other public bodies that determine whether research meets ethical requirements

Secondary use of data: studying data for another purpose for which it was originally intended

Survey data: information collected directly from and about individuals or groups

Surveillance data: data about individuals in the context of everyday life, e.g. CCTV video capture, Fitbit, web activity, GPS tracking) collected for the purpose of tracking or influencing those observed



