DATA GOVERNANCE AND MANAGEMENT TOOLKIT
AND LEARNING SERIES

Webinar #5: Acquiring & Working With Data

Series Facilitator: Harmony Johnson sɛƛakəs

March 5, 2021

WHAT ARE WE DOING HERE?

- Organized by self-governing Indigenous Governments (IGs) to provide practical knowledge about data to support Indigenous Governments in their responsibilities

HOUSEKEEPING

- This webinar will be recorded and posted (in part or in full) publicly to the website – however, you will not be recorded
- Please turn on your video and wave hi
- To ask questions…
  - Use the chat OR
  - Use the hand up function

Some personal notes:

- I apologize in advance if I use the term ‘Nation’ instead of ‘Government’
- I will often be looking at my second monitor or my notes
- I will try not to talk too fast
- We will do our best with the technology and may make some mistakes along the way

ZOOM

Mute/unmute

Start/stop video

Open the chat

Raise your hand

KEY TERMS FOR TODAY (*SEE TOOLKIT GLOSSARY)

- **Data:** Qualitative or quantitative information that is collected for reference or analysis, and includes stories, facts, measurements, values, qualities, or observations. The collection, analyses and interpretation of data and information is a basic activity of government.

The power of data

- Improve outcomes
- Inform policy, practice, planning and decision-making
- Communicate to citizens
- Fulfill reporting requirements
- Evaluate programs and measure quality
- Celebrate success
- Contribute new knowledge and understandings

KEY TERMS FOR TODAY (*SEE TOOLKIT GLOSSARY)

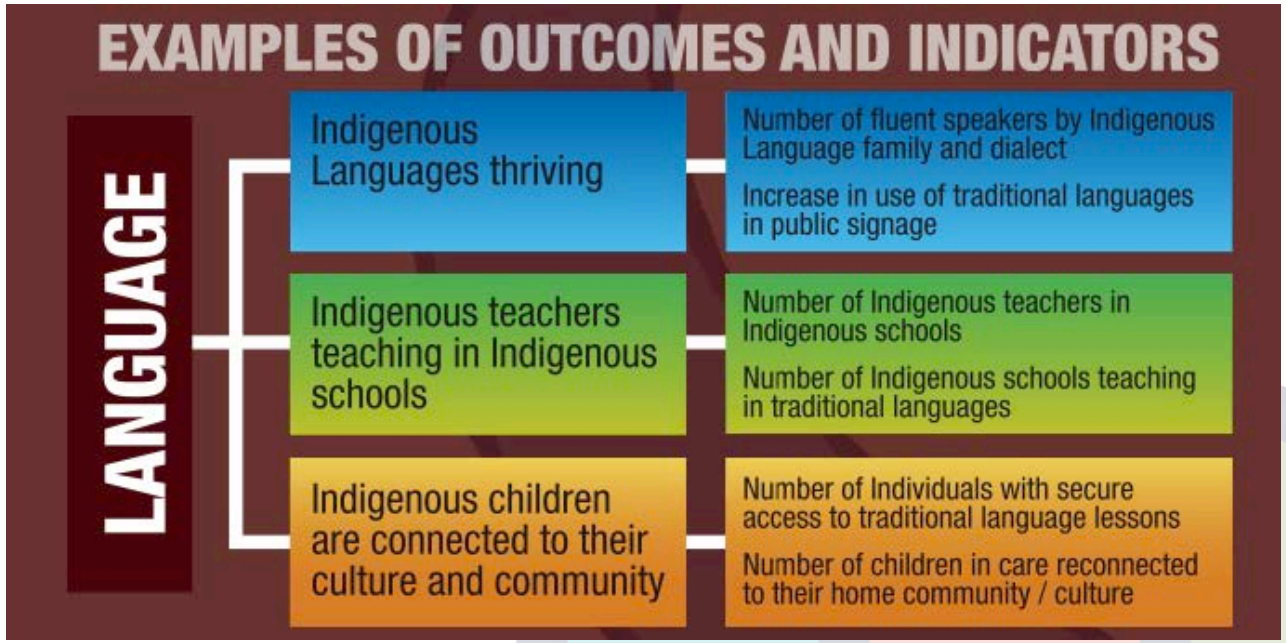
- **Administrative data:** Data collected and used for the purposes of performing governmental operational functions. This includes data collected and used during delivery of a program or service and data generated and used for internal government processes, such as Human Resources and Finances. Administrative data may be generated through transactions, record keeping, or registration, amongst other activities. Administrative data may be used secondarily for analytical purposes, but its primary function is to serve an operational function.

- **Primary data:** Data that has been collected for the express purpose of a project.

- **Secondary data:** Data that has been collected for some purpose other than the project it is currently being used for.

KEY TERMS, CONT.

- **Outcome:** Measurable statements that let you know when you have reached your goals.
- **Indicator:** A piece of data or a composite of several pieces of data that, tracked over time, measure outcomes.
- **Performance Measurement**: The process of collecting, analyzing and/or reporting information regarding the performance of an individual, group, organization, system or component towards intended results.



SELF-ASSESSMENT TOOL

Department-Level/Administrative Data: Does your government use socioeconomic data from different departments effectively in decision-making? Consider data that might be available in hard copy records and database systems. For example, your financial system might have information about post-secondary support, your program staff may collect data about program participation rates, etc.	*The questions in this section will help you conduct an environmental scan of the internal socioeconomic data your government already has compiled. You need a clear picture of the current data landscape within your SGIG in order to assess needs and gaps.*			Acquiring and Working with Data > Internal Data Sources and Primary Data Collection > Administrative Data
Internal Surveys: Has your government conducted surveys related to socioeconomic outcomes? (For example: health, language and culture, housing, education, employment)				Acquiring and Working with Data > Internal Data Sources and Primary Data Collection > Primary Data Collection

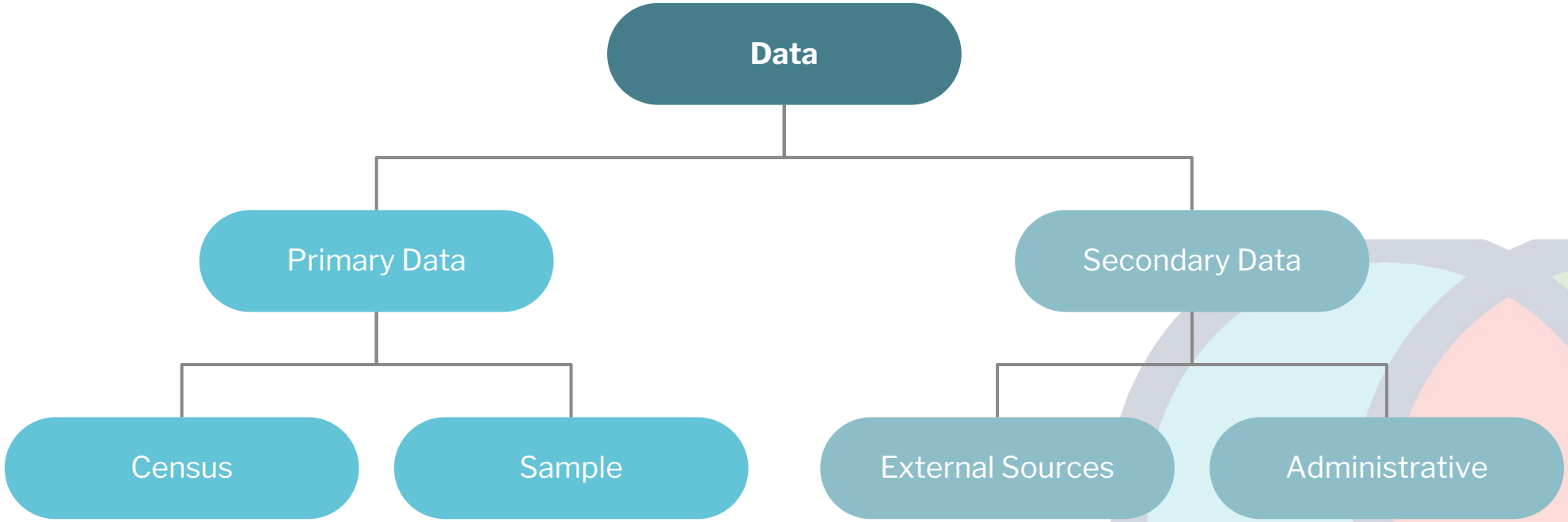
DATA GOVERNANCE AND
MANAGEMENT TOOLKIT
AND LEARNING SERIES
Acquiring & Working With Data
Hannes Edinger, Big River Analytics

Webinar Outline

1. Key Terms
2. Motivation
3. Three Key Takeaways
4. Understanding and Stating Objectives
5. External/Existing Data Sources
6. Administrative Data
7. Primary Data Collection

Key Terms

Relevant data types



Motivation

Many practical problems are best solved using a data-oriented approach, for example:

Informing indicators *(why we're here!)*
CASE STUDY: Housing work

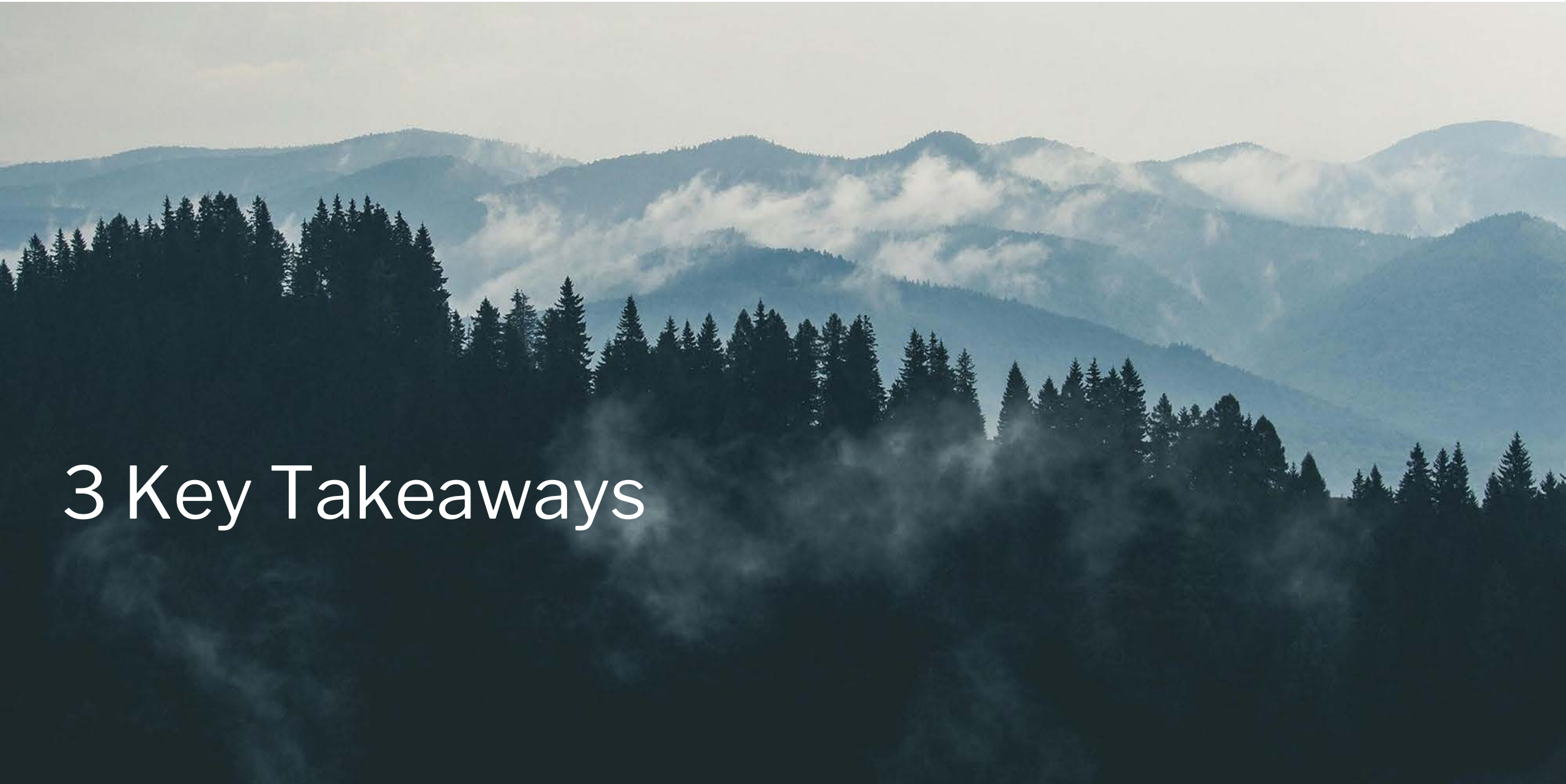
Providing a rationale for action *(motivating others)*
CASE STUDY: ISETS agreements

Measuring performance *(telling the story of your success)*
CASE STUDY: Community-based reporting

Motivation

This is hard work, but the payoff is worth it

- Achieving a balance between internal and external resources is key
- The success we've seen working with Indigenous governments that have become expert data users is tremendous



3 Key Takeaways

TAKEAWAY 1 : Understanding Your Information Needs

"~~The most important step~~ in acquiring data is deeply understanding your information needs. Developing this understanding is best achieved in a structured, coordinated setting like a Statement of Objectives meeting."

TAKEAWAY 2: Novice to Expert Data Users

Novice Data Users

- Rely too much or too little on external support
- Haven't experienced the benefits of effective data use
- Are too bold (Survey Monkey!) or too fearful of a data-related project (status-quo bias)
- No structure around info needs
- Lack framework for how to conduct this type of work

Expert Data Users

- Have some internal capacity
- Can articulate info needs
- Have done (or are open to) organizational change work to support data-related initiatives
- Have found the sweet spot for external support, have a trusted network of service providers
- Have experienced the benefits of effective data use

TAKEAWAY 3: Primary Data Collection

A bit like a liver transplant - hard, uncomfortable, and a serious undertaking but sometimes you don't have another option

Key is to understand when it's necessary and worth it (you don't do a liver transplant when you've got a sore stomach)

When it's required, the benefits can be tremendous!

- After you've gone through a proper primary data collection exercise you will moved along the data use continuum towards becoming an expert

  These exercises can serve as a catalyst for organizational change more broadly

External & Administrative Data Sources

External Data Sources

- Statistics Canada
  - Census
  - Aboriginal Peoples Survey
  - Health Surveys
  - Many more to be used as a *comparison*
- FNIGC
- Vital Statistics
- Provincial / Territorial Statistical Agencies

Statistics Canada

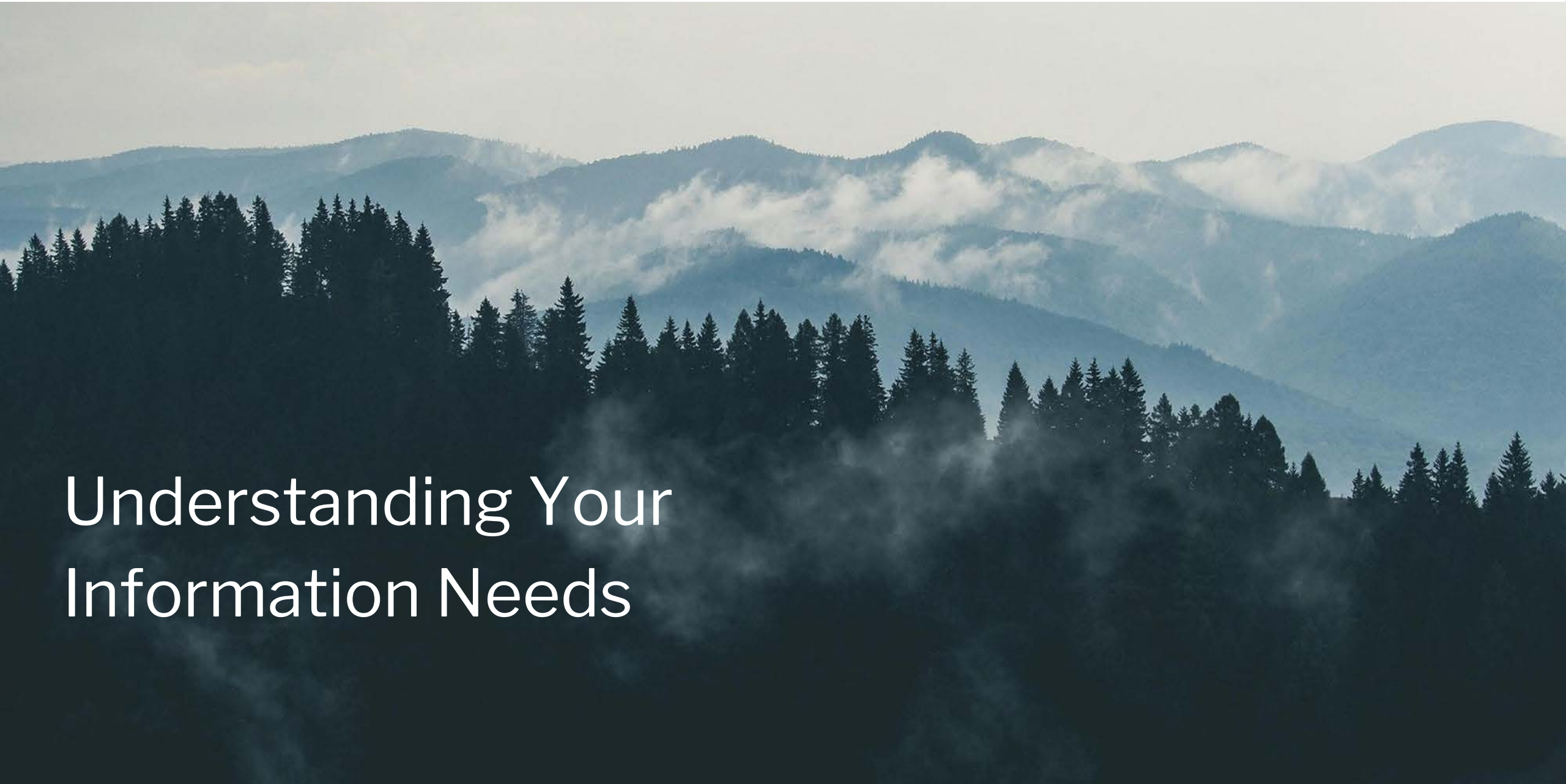
- Census
  - Most comprehensive data set
  - Standard geographical classification
  - Opportunity to do a linking exercise
- APS
  - Off-territory
  - Small sample
  - Designed for reporting at provincial level
- RTRA

Accessing the Census

- Website
- Download questionnaire
- Custom tabulations
- Research Data Centres

Administrative Data Sources

- Accounting or program-related software
  - Number of post-secondary students receiving assistance
  - Number of citizens with harvesting permits
- Citizenship/membership database
  - Number of citizens/members
  - Demographic information
- Permit database
  - Industry activity on-territory
- Consider potential other data streams! Simple, modern databases can be low-cost and very helpful



Understanding Your Information Needs

Overview

- What is ‘the’ Statement of Objectives process
- Steps to Developing a Statement of Objectives
  a. Statement of Information Needs
  b. Data and Data Users
  c. Concepts and Operational Definitions
  d. Research Design
  e. The Analysis Plan

What is ‘the’ Statement of Objectives Process

Process where you set out to answer key questions to inform your data project. E.g.:

- What information is of highest use/importance
- What information is already collected or available
- What survey responses will be used for
- What definitions, terms, and considerations are important

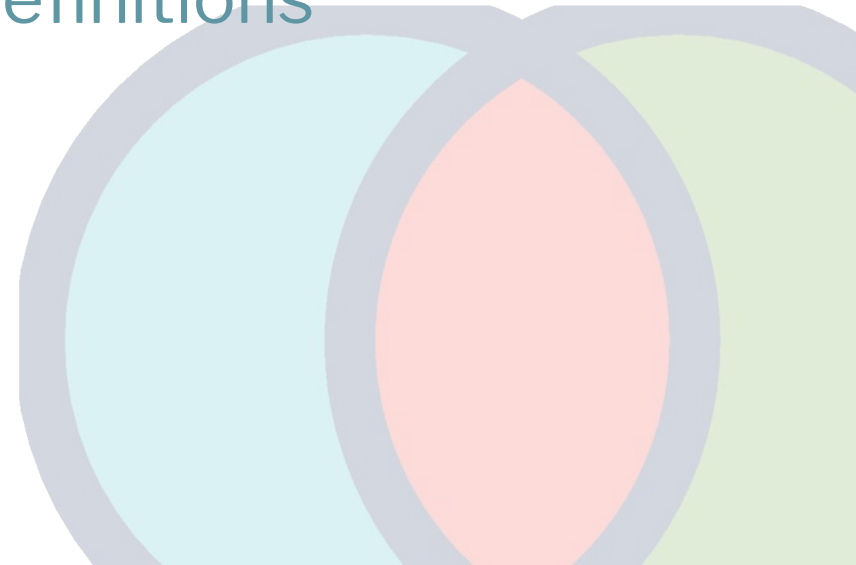
What is ‘the’ Statement of Objectives Process

The key aim is to ensure your data project actually:

- Meets the objectives you want to achieve
- Answers the key research questions you set out to achieve

Steps

1. Statement of Information Needs
2. Data and Data Users
3. Concepts and Operational Definitions
4. Research Design
5. Analysis Plan



Statement of Information Needs

- Why has this project been suggested?
- What are the stated and unstated goals of those involved?
- What questions are you interested in answering?
- Are there specific topics you would like to better understand?

Statement of Information Needs

Ways to move from Research Interests to Research Questions

Initial Research Interests

The more specific your research question, the easier it will be to define your target population and the questions necessary to create a successful survey.

Literature Review

Conducting a literature review establishes what contribution your work can add to the current research and helps to refine your research question further.

Secondary Data

Secondary data analysis highlights what questions you can and cannot answer without primary data collection. Secondary data can also be used to supplement findings or draw comparisons.

Research Question

Coming to a 'final' research question can involve several iterations. Your question needs to be sufficiently well specified, practical, and measurable.

Data Users and Uses

- Who are the main users?
- What will the information be used for?
- Are you analyzing a relationship or describing a situation?
- What types of decisions will be made based on the data, and what might be the consequences of these decisions?
- Who is the audience of your final output?

Concepts and Operational Definitions

- Who are you interested in (target population)?
- Where and what are the units (individuals, geographies) of interest?
- What is the reference period and is there a reference group?

    Is valid comparison across years important?

    Short, medium, and long term outcomes?

Research Design

Dependent on your project

- E.g. Questionnaire design & survey strategy
- E.g. Methodology

The Analysis Plan

- What tables, graphs, tabs, & charts would be of highest value?
- Does the design of your research project actually answer what you set out to understand?

For consideration:

Are there relevant comparison populations?

What context is important for understanding this result?

Does the presentation of your data/analysis articulate the most information?

From Data to Output: The Analysis Plan

Survey Question

Responses

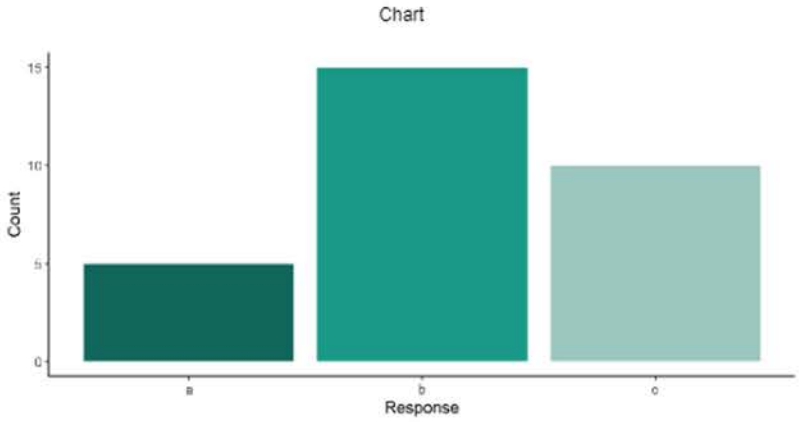
response
a
b
c
a
b
b
c
b
c
b

Analysis/Processing

Processed Data

response	count
a	2
b	5
c	3

Plotting



Primary Data Collection

Questionnaire Basics

General Rules of Thumb

- Surveys should aim to take less than 5 minutes (15 questions) to complete
- Where possible, use images, Likert Scale questions, randomization, and limit qualitative questions
- Goals for Survey Design include:
  - Be as streamlined as possible - no unnecessary questions
  - Does not introduce bias
  - Be as user friendly/easy for respondents as possible
  - Uses a logical survey flow
  - Acknowledge your respondents are adults

General Rules of Thumb

- Every survey has an **attention budget**

  When that budget is spent, respondents will leave early or answer haphazardly

  Each question costs attention; more complex questions cost more

- **Consistency** comes first

  Use the same term for something throughout a survey

  e.g. Don't use “climate change” and “global warming” interchangeably

General Rules of Thumb

- **Avoid negatives**
  - The word *NOT* is easy to overlook even when highlighted, and processing negatives costs attention

- Be **exact**

  e.g. Use terms like “in the last 7 days” instead of “in a typical week”

- **Pre-empt guessing**

  e.g. Use bins like “Between $20,000 and $30,000” for large quantities that people are unlikely to know exactly

General Rules of Thumb

- **Neutral, unassuming language** gets honest responses.
  e.g. Ask about “small dogs”, not “small, yappy dogs”.

- **Use skip patterns** to justify assumptions
  e.g. Ask “Do you use any time management software?”, and if the answer is no, skip over “What time management software do you use?”

General Rules of Thumb

- **Allow non-responses** like “Don't know” or “Prefer not to answer”
  Missing a value for one variable is much better than missing half a survey to frustrated dropout

- Pre-made **Likert scales** for questions about degrees of agreement or importance are readily available online

Test your Survey Overview

Start by taking the survey yourself:

- Can you answer all the questions easily and accurately?
- Could your friends, family, peers, colleagues answer all the questions?
- Could people with very different life experiences answer the same questions easily and accurately?

Test your Survey Overview

Think about **all the possible types of individuals** who might respond to your survey (e.g. asking questions of students' experiences → graduate from 3-month online college course, current PhD student, doing practicum in nursing etc.)

- Could they answer each question? Are questions confusing given their circumstance?
- What would their answer mean given their situation or context?
- Do they have the necessary information to respond?

Test your Survey Overview

Think bigger picture:

Length, energy/thought required, relevance to your research question(s), ability to generate tables/chart in your analysis plan, aesthetic, legibility, flow, longevity

Think smaller picture:

Font size and colour, spelling

Individual question structure - do's and don'ts and checklist

Test your Survey Checklist

Review each question from the perspective of diverse participants and consider the following:

Necessity

- ❑ Is the question necessary for the success of your analysis plan or survey strategy?
- ❑ Does the question generate data that can be processed, analyzed, and summarized?
- ❑ Is this question answered elsewhere in the survey or secondary data?

Test your Survey Checklist

Clarity

- ❑ Do you have to read the question more than once or twice?
- ❑ Did you use spell check?
- ❑ Does the location in which the question appears in the survey make sense (logic and relate to surrounding content)?
- ❑ Is the question leading?
- ❑ Are your forced into an answer you might disagree with?
- ❑ Does the question use absolutes (always, never, all, none)?
- ❑ Does the question have negatives (not, none, never)?
- ❑ Is it clear what time frame is relevant for your answer?

Test your Survey Checklist

Simplicity

- ❑ Can you remove filler or unnecessary words? (Keep it brief)
- ❑ Could a grade 8? Grade 6? Grade 4? Level reader understand your question?
- ❑ Are all the words and concepts defined or common place?
- ❑ Is the question asking one thing?
- ❑ Does your question have mutually exclusive answer sets where it shouldn't? (i.e., Are you a student or employed?)

Test your Survey: Pilot and Review

Conduct a pilot survey (have (lots, if possible,) people take your survey)

- Go through the survey yourself
- Pilot / Test the survey with actual respondents
- Pilot / Test the survey with other colleagues
- Review your results and generate output
  - Is it as expected? Is it suitable for the desired analysis plan? Do you need to add or refine questions to get desired output

Test your Survey: Pilot and Review

Additional review

Seek targetted feedback on the survey from stakeholders (e.g. other departments & colleagues, focus groups/pilot with Citizens or members but avoid the kitchen sink (focus review not on adding questions, but the legibility and effectiveness of the questions included)

Monitor data collection

Upon release, again, check the validity and quality of responses for anything unexpected (treat your first respondents as a pilot if required)

Identify successful language/channels to improve response rates

Primary Data Collection

Survey Design

Examples: Census vs Survey

Census	Survey
Statistics Canada 5-year census of all Canadian residents	Statistics Canada survey on medical needs sent out to 1000 random residents
Enumeration of all real estate properties within the limits of a city	Random selection of 500 properties with the limits of a city
Tagging and tracking of every animal of a critically endangered species	Capture, tagging, and recapture of a common species

Convenience Sample

EXAMPLE 1

Getting a count of the local wildlife by driving around near the edges of town.

- Much easier, cheaper, and faster than flying over randomly selected areas in the region and counting from overhead.
- Sample will be almost entirely populated by animals more accustomed to humans.

Convenience Sample

EXAMPLE 2

Phoning randomly selected people for a voluntary survey

Most realistic scenario for any sort of opinion survey.

Without adjustments, sample will be over representative of people with the time and inclination to answer your survey.

Will not represent people who never answer their phones.

Can be made more representative with adjustments for known variables like age and sex.

Simple Random Sample (SRS)

- A type of probability sampling, a simple random sample is a method where **everyone has the same probability of being selected.**
- Meant for homogeneous populations that are equally reachable.

  e.g. a survey from a cohort of students with similar background and motivation.

  e.g. products from an assembly line.

Simple Random Sample (SRS)

Strengths

Simple to implement, analyze, and understand.

Population members are all assigned to the sample in one step.

Weaknesses

Limited range of use cases. If our population is non-homogeneous, some members may be harder to include the sample than others, and an SRS will give biased results.

Stratified Sample

- A type of probability sampling, a stratified sample is a method where a **fixed number of individuals are selected from each stratum, or "group", within a population.**
- Strata should be something with a previously known distribution.
    - e.g. age or sex in population of people.
    - e.g. different flavours of product coming off an assembly line.

Stratified Sample

Strengths

Wider use of cases

Accounts for some types of under- and over- representation

Weaknesses

Requires additional setup and planning

Post-Stratification

- Post-stratification is a technique you can use if you are unable to reach your target population.
- It is useful to correct for potential response bias or bias introduced by the way you've sampled

Post-Stratification

- To understand **response bias**:
  - Consider a scenario where you are collecting data on students in post-secondary and only those who are in engineering complete your survey

- To understand **post-stratification for convenience-based samples**:
  - Consider a scenario where you are collecting data on students in post-secondary and only students over the age of 40 complete your survey

Post-Stratification

- The goal of using survey weights in these context is to **adjust the relative importance (weight) of some respondents in an effort to mitigate the impact of bias** in our estimates due to non-random sampling, or low response rates from particular subsets of the population.
- You can test whether you might need to apply survey weights by comparing characteristics of your respondents with known characteristics of the target population.
- Ideally, you can change your survey strategy to avoid needing this technique, but this is a useful approach

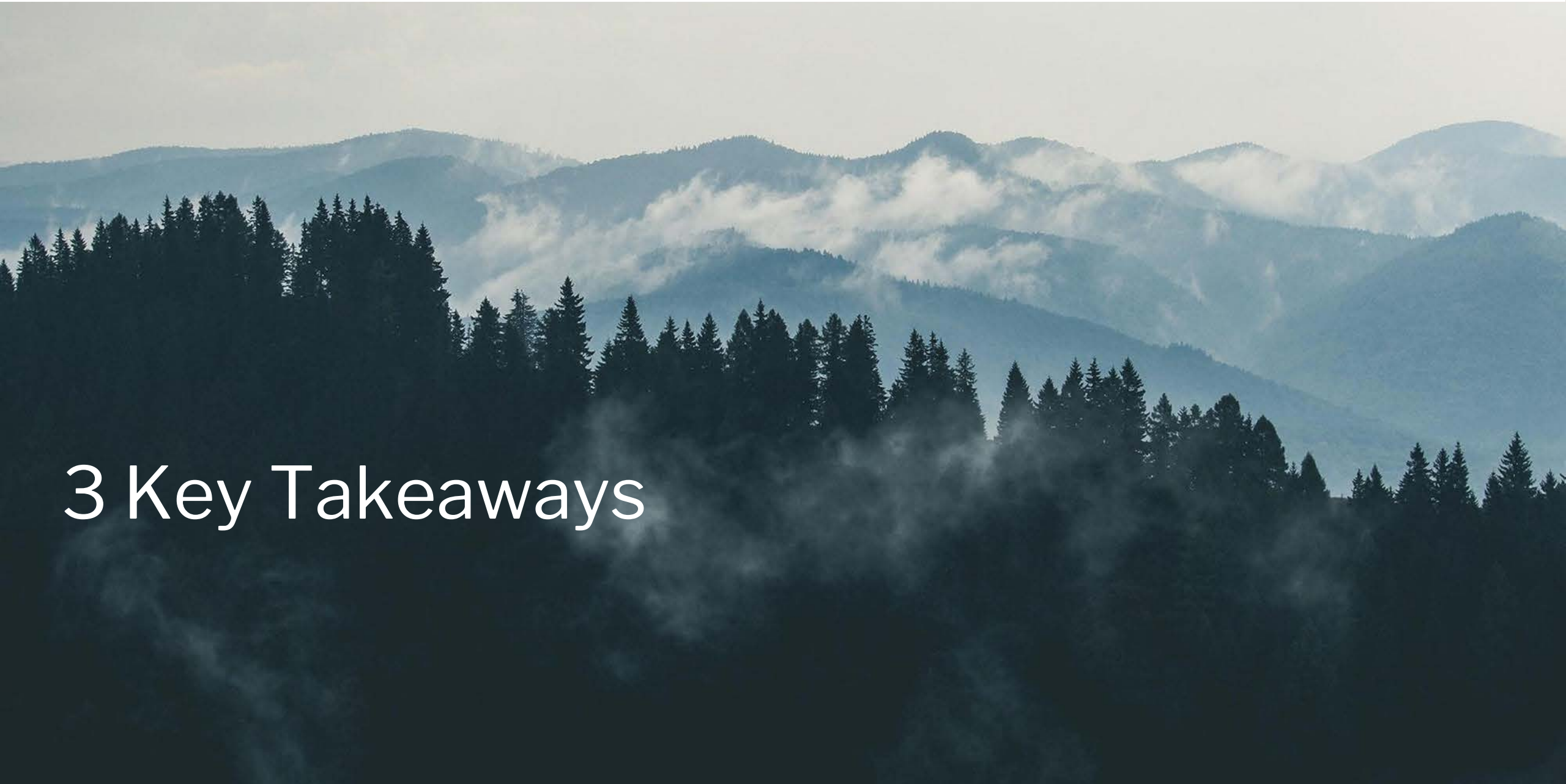
Privacy

- All data collected through a survey should be considered potentially sensitive and we should respect survey respondents willingness to participate. Consider how to:
- Ensure transparency, openness, and choice around data use, that is,
- Respondents understand what information is being collected, what it will be used for, what risks there may be in sharing data, any data retention or privacy policies they are subject to in clear and plain terms
- Respondents should be able to opt out of particular uses of their information with no penalty

Privacy

Minimize potential risks, that is,

- Analysts should limit collecting personal information unless it is necessary
- Analysts/organizations should limit access to data with personal information (i.e., anonymize responses, store appropriately)
- When private information is collected with the intent to share aggregate data, analysts/organizations should carefully consider whether individual information could still be identified
- When third-party services, i.e. Survey Platform Software, is used, remove all data and necessary documentation at the conclusion of the survey, and storing it on a secure, backed up, first-party network drive



3 Key Takeaways

TAKEAWAY 1 : Understanding Your Information Needs

"~~The most important step~~ in acquiring data is deeply understanding your information needs. Developing this understanding is best achieved in a structured, coordinated setting like a Statement of Objectives meeting."

TAKEAWAY 2: Novice to Expert Data Users

Novice Data Users

- Rely too much or too little on external support
- Haven't experienced the benefits of effective data use
- Are too bold (Survey Monkey!) or too fearful of a data-related project (status-quo bias)
- No structure around info needs
- Lack framework for how to conduct this type of work

Expert Data Users

- Have some internal capacity
- Can articulate info needs
- Have done (or are open to) organizational change work to support data-related initiatives
- Have found the sweet spot for external support, have a trusted network of service providers
- Have experienced the benefits of effective data use

TAKEAWAY 3: Primary Data Collection

A bit like a liver transplant - hard, uncomfortable, and a serious undertaking but sometimes you don't have another option

Key is to understand when it's necessary and worth it (you don't do a liver transplant when you've got a sore stomach)

When it's required, the benefits can be tremendous!

- After you've gone through a proper primary data collection exercise you will moved along the data use continuum towards becoming an expert

  These exercises can serve as a catalyst for organizational change more broadly

CLOSING

- Next Webinar March 19
  - Session 6: Data Management and Information Management
- Utilize the advisors to support your work – contact Mieke at mcoppes@Nvisiongroup.ca
- Please briefly complete the pop-up poll, *"How would you rate the usefulness of this webinar?"*
- For more information: https://indigenousdatatoolkit.ca

Thank you!

Questions?