Open Data Sharing for Populations at Risk of Re-Identification

Elizabeth G. Condliffe PhD, MD, FRCPC PI with a Problem
Stacey Page PhD, Chair of CHREB
Amanda Rande MSc Worker Bee Finding Solutions

PONI
UNIVERSITY OF CALGARY

What is Open Data Sharing?

access

re-use

re-distribution

transparency

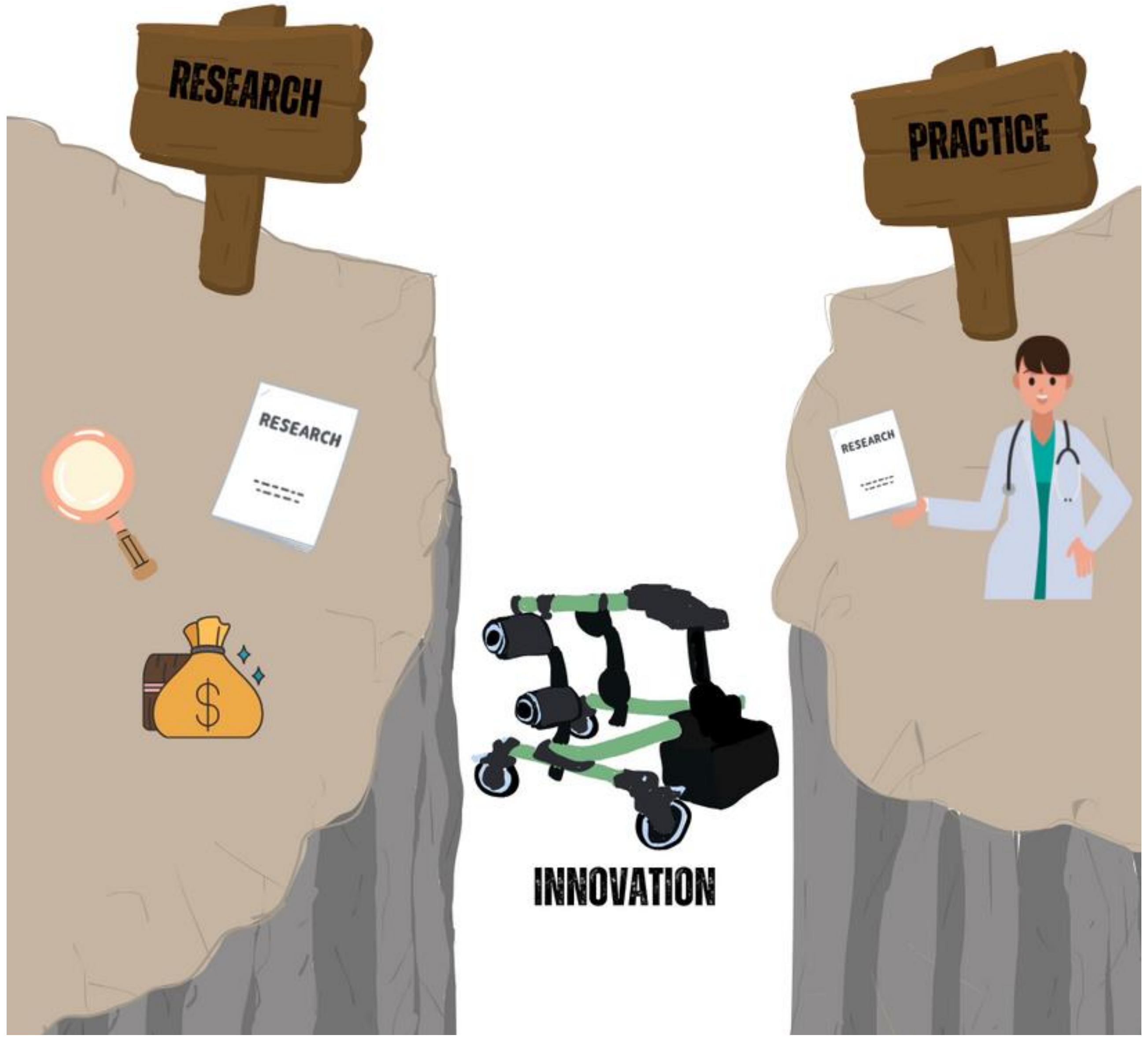
collaboration

enables researchers, policymakers, and the public to utilize data

PONI

UNIVERSITY OF CALGARY

Research-to-Practice Gap



- Time to prove efficacy
- Time to figure out implementation
- Time to prove effectiveness
- High cost $$$

Our Participants

Research Barrier	Open Data Sharing Solution
studies on rare diseases are often underpowered	build upon other research to create meaningful evidence
novel technologies are at risk of the research to practice gap	share data and results faster and with a broader community
parents and families unaware of research outcomes	data and research outputs shared more broadly

but...

explaining open data sharing and ensuring privacy can be a challenge

Objectives

By the end of this lunch and learn session you will be able to:

- Describe open data sharing to colleagues, potential participants, and proxies
- Evaluate the level of risk of re-identification for data
- Discuss approaches for ethical, legally sound, open or partially open data sharing

Research Ethics Fundamentals

- TCPS2:
  - Respect for persons, concern for welfare, justice
  - Consent (Chapter 3), fully informed, given voluntarily, given by individual with capacity
    - Broad consent
  - Privacy and confidentiality (Chapter 5)
    - Secondary use, consent and identifiable/deidentified
  - Justice (Chapter 4)

Broad Consent - TCPS2 Article 3.13

- is consent for unspecified, future research (i.e., storage, secondary use).
- broad consent always includes specific restrictions

Relevant info for participants:

- What is being collected and stored for reuse and why
- Voluntariness, ongoing consent, options for withdrawal (if any)
- Risks and benefits of storage and use in unspecified future research
- Information about the repository and its governance

Consent and future use

- Can participant consent be mandatory for future, unspecified research use of *de-identified* data?
  - it is strongly encouraged that this be a choice
    - risk of coercion
    - Inclusion/exclusion and justice
- CHREB template language

It is your choice whether or not to let researchers share your de-identified data for research in the future. If you say “yes,” you can change your mind later. While data used to that point cannot be withdrawn, it will not be used in further research. If you say “no,” you can still fully participate in this study.

Please indicate your choice:

- ______YES, use my data in other research studies
- ______NO, do not use my data in other research studies

Deposit of existing data where consent was not originally sought for secondary use

https://ethics.gc.ca/eng/depositing_depots.html

- Preferences regarding deposit and recontact unknown
    - *Must contact and seek consent to deposit, subject to REB approval*

- Preferences unknown regarding deposit or recontact, seeking consent impracticable
    - *OK to deposit, subject to REB approval*

- No known preference regarding deposit, known refusal of recontact
    - *Must not recontact, generally must not deposit data*
    - *Exceptions – exceptional contexts, anonymized non-sensitive data, subject to REB approval*

Depositing Existing Data in Public Repositories

REB to consider

- if participants would have consented to deposit and sharing
- if the researchers’ proposals will continue to respect the terms of consent
- adequacy of consent waiver justifications
- data sources (participants, other researchers, external sources)
- type of data and sensitivity
- level of identifiability/privacy measures
- whether a unique community or group is involved
- quality of the repositories (governance, policies related to data access and reuse, and safety, security and protections in place)

Traditional Research Pathway

Traditional research outputs:

- manuscripts
- infographics
- typically not translated into ‘lay’ results

“Data is available upon reasonable request to the corresponding author”

Research Pathway with Data Sharing

Data Collection

Participant ID

Data Extraction

Data Analysis

De-Identification

Research Outputs Created

Data is Securely Stored

ID Re-coded

Quasi Identifiers Removed

Data Deposited

Data is re-used

What is re-identification

Directly identifying

Quasi-identifier

Anonymized

Name

Personal Health Number

Address

Telephone Number

Email Address

Driver’s License Number

Bibliometric Data

Age

Diagnosis

Gender

Sex

Postal Code

Approximate Location

Data Matching

Sex + Age = 18/1,000

Sex + Age + Primary Diagnosis Code = 901/1,000

Common Disease (Diabetes)

Sex + Age + Postal Code = 92.5%

Sex + Age + Forward Sortation Area = 1.7%

Rare Disease (Scabies)

Sex + Age + Postal Code = 100%

Sex + Age + Forward Sortation Area = 66.7%

Likelihood of Identifying Individuals in De-Identified Record-Level Data. *Public Health Surveillance Bulletin*, Bulletin No.2, June 2013

Evaluating Risk

Assessing the risk of re-identification requires evaluating the dataset’s content, context, and the likelihood of external data sources converging with it.

Example:

100 participants used a novel rehabilitation technology

The dataset contains:

- Age
- Sex
- General diagnosis

Only one participant is over the age of 18, this participant appeared in a news article and posted on their social media about using the device

Evaluating Risk

Assessing the risk of re-identification requires evaluating the dataset’s content, context, and the likelihood of external data sources converging with it.

Example:

100 participants used a novel rehabilitation technology

The dataset contains:

- Age
- Sex
- General diagnosis

Only one participant is over the age of 18, this participant appeared in a news article and posted on their social media about using the device

Content + Context = **High** Risk of re-identification

Approaches to Data Sharing

Fully Open	**Semi-Public/Registered**
unrestricted access to datasets	require approvals
commercial and non-commercial access	access determined by research team & ethics board
guest book to track who access the data	data transfer agreements and/or terms of use
limited control of what happens to the data	control of what happens to the data

Our Approach

Why not do both?

Fully Open	**Semi-Public/Registered**
aggregate demographics	granular demographics
age bands	age (years)
general diagnosis	more specific diagnostic information
summarized or aggregate qualitative data	more detailed qualitative data

engagement with patient partners and the ethics throughout the process ensures we are sharing data in the way our partners want while meeting the ethical and legal requirements

Thank you!

Thank you to the HBI for funding our open science initiatives

HOTCHKISS BRAIN INSTITUTE

PONI

UNIVERSITY OF CALGARY