NSERC ResNet Data Policy

John Clark

NSERC ResNet
Monthly Network Meeting
2021-10-15

Agenda

- Motivation
  - User Survey
  - Open Science
- Data Policy Components
  - Roles and Responsibilities
  - Data Publication
  - Exceptions
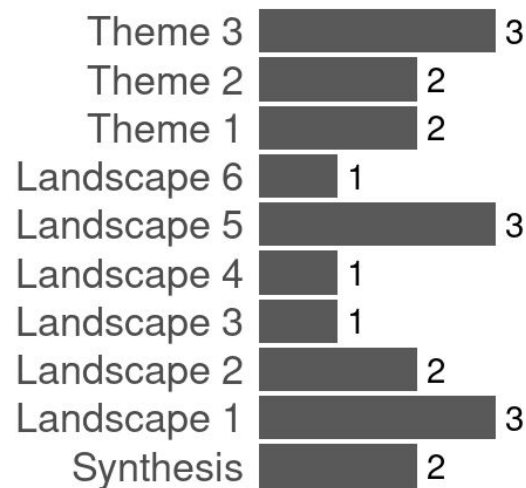- Examples
- Next Steps/Discussion

Motivation

Data User Survey

- Conducted June 2021
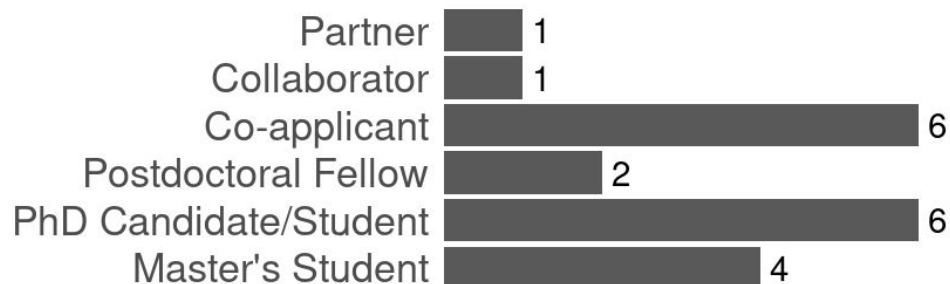- Identify network needs
- Small sample size (n=20)

Affiliation
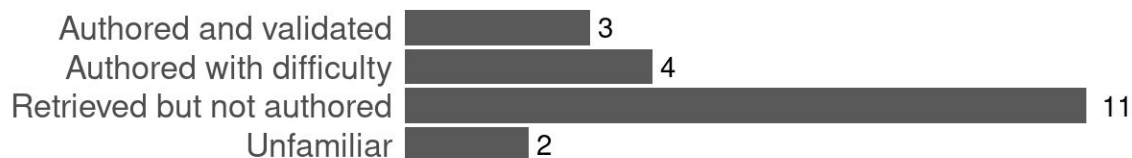
What is your affiliation with ResNet?

Theme 3	3
Theme 2	2
Theme 1	2
Landscape 6	1
Landscape 5	3
Landscape 4	1
Landscape 3	1
Landscape 2	2
Landscape 1	3
Synthesis	2

Role

What is your role in ResNet?

Partner	1
Collaborator	1
Co-applicant	6
Postdoctoral Fellow	2
PhD Candidate/Student	6
Master's Student	4

Data User Survey

Metadata
What is your experience with metadata?

Authored and validated	3
Authored with difficulty	4
Retrieved but not authored	11
Unfamiliar	2

Spatial Data
Is spatial data important to your research?

Very	11
Signficantly	4
Somewhat	4
Not at all	1
Not sure	
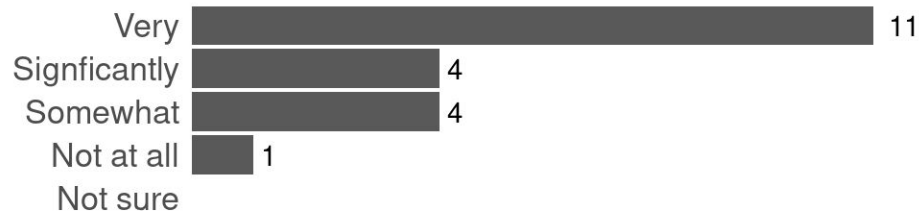
Sensitive Data
Does your research involve sensitive data?

Yes	10
No	8
Not sure	2

Data User Survey

Data Publishing Policies

Does your lab/institution/funder have policies in place for publishing data?

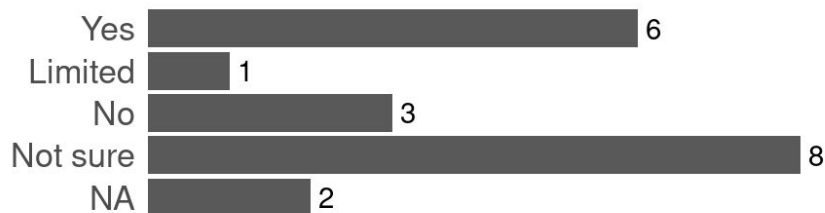|	
Yes	10
Informal	1
No	2
Not sure	7

Data Management Plans

Does your lab/institution have a Data Management Plan?

Yes	3
Partial	2
No	7
Not sure	8

Sensitive Data Policies

Does your lab/institution have a policy for handling sensitive data?

Yes	6
Limited	1
No	3
Not sure	8
NA	2

Data User Survey

Storymap Interest

Are you interested in communicating your research through a storymap?

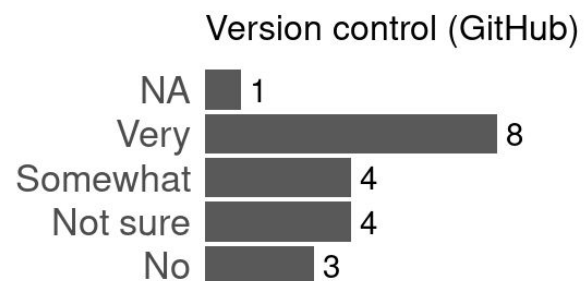|	
Very	12
Somewhat	4
No	
Not sure	4
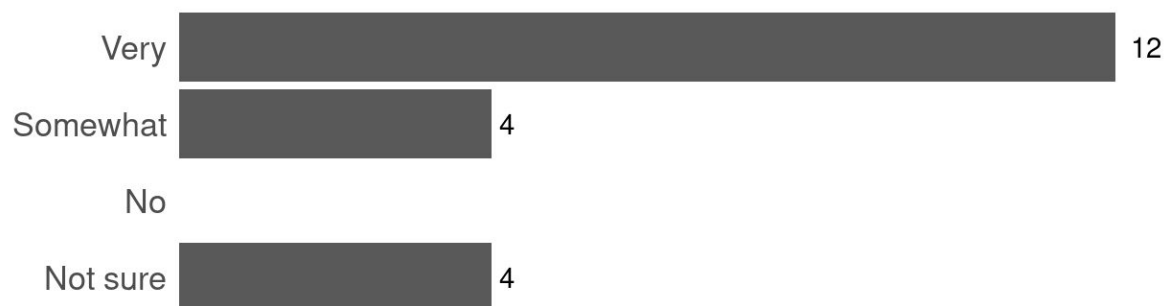
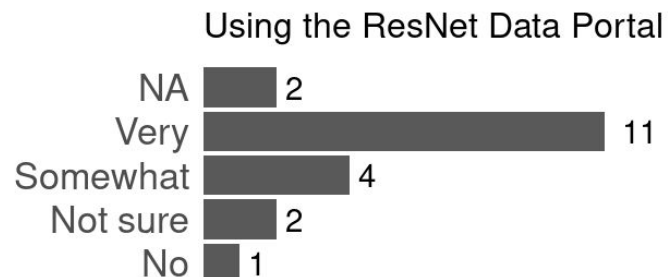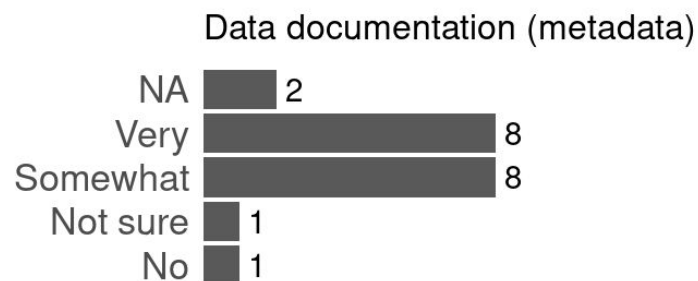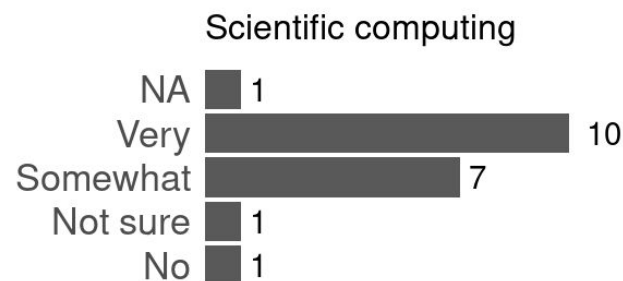## Dashboard Interest

Are you interested in communicating your research through a dashboard or other interactive tools?

Very	11
Somewhat	9
No	
Not sure	

Data User Survey

Are you interested in training, workshop, or round-table sessions for these topics?
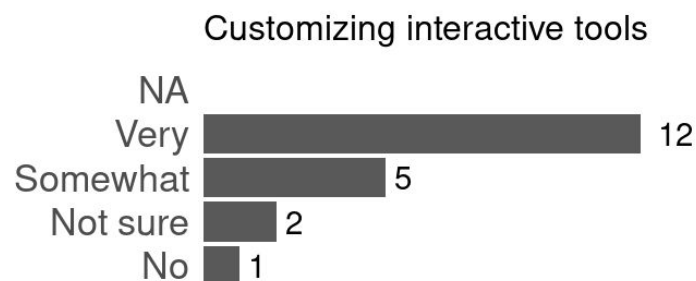
Scientific computing

NA	1
Very	10
Somewhat	7
Not sure	1
No	1

Data documentation (metadata)

NA	2
Very	8
Somewhat	8
Not sure	1
No	1

Data processing

NA	2
Very	9
Somewhat	7
Not sure	2
No	

Using the ResNet Data Portal

NA	2
Very	11
Somewhat	4
Not sure	2
No	1

Version control (GitHub)

NA	1
Very	8
Somewhat	4
Not sure	4
No	3

Customizing interactive tools

NA	
Very	12
Somewhat	5
Not sure	2
No	1

Data User Survey

Takeaways

- We rely on metadata but don't always create it
- Spatial data is common but not universal
- Sensitive data is a frequent concern
- Existing/external policies are inconsistent
- Excited to incorporate research into data driven apps

Data User Survey

Takeaways

- We rely on metadata but don't always create it
- Spatial data is common but not universal
- Sensitive data is a frequent concern
- Existing/external policies are inconsistent
- Excited to incorporate research into data driven apps

Therefore, we need a **Data Policy**

- A document formalizing data publication and use
  - Establishes standards
  - Fosters participation
  - Accommodates different types of data
  - Respects restrictions on sharing
- Data publication facilitates data driven apps
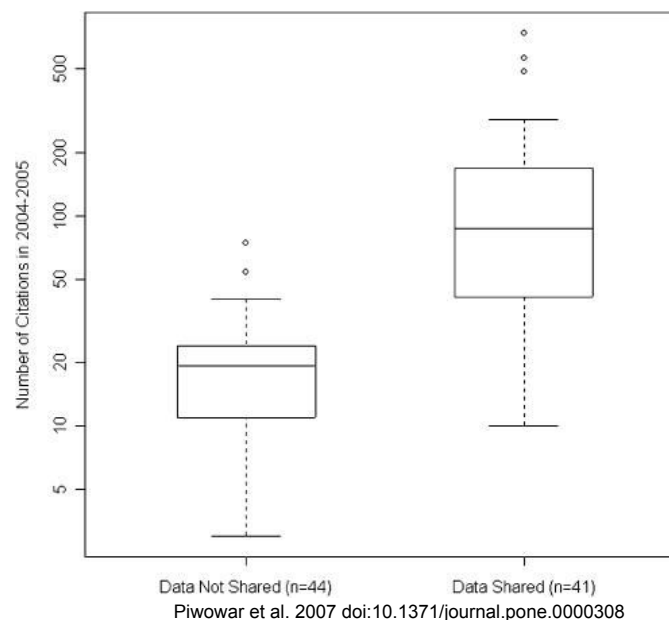

DATA POLICY

Data Publication: What?

- Process of preparing and releasing research data for use by others
- FAIR Data Principles
  - **F**indable
    - Metadata and data should be easy to find for both humans and computers
  - **A**ccessible
    - Must be able to retrieve after finding
  - **I**nteroperable
    - Data need to interoperate with applications or workflows for analysis, storage, and processing
  - **R**eusable
    - Optimise the reuse of data. Metadata and data should be well-described so that they can be replicated and/or combined in different settings
- Requires active effort to implement

Data Publication: Why

- ● For the network
  - ○ Collaboration is the heart of our network
  - ○ Make connections across Landscapes and Themes
  - ○ Shared repository of knowledge
  - ○ Standardized for interoperability
- ● For everyone
  - ○ Reproducibility, transparency
  - ○ Reduce duplication of effort
  - ○ Public interest
- ● For you
  - ○ Increasingly required by funder/publisher
  - ○ Enhance your project management
  - ○ Share data for broader use
  - ○ Incorporate into data driven apps
  - ○ Amplify the impact of your work
    - ■ Published data is citable data

Piwowar et al. 2007 doi:10.1371/journal.pone.0000308

Feedback: Motivations

Have you used open access data in your work? What from where? How was the experience?

Are there additional obstacles and incentives to implementing open science practices?

https://bit.ly/ResNetDataFeedback

Policy Components

Data Policy

What's in it?

- Defines roles and responsibilities, data publication, data usage, and governance
- Think of it as the "Why, What, Who, When" for open science within NSERC ResNet

What isn't?

- Generally, the "Hows"
- Two separate documents in development complement Data Policy
  - **Data Guides**
    - Series of instructions and tutorials to assist researchers with data management
    - Detailed workflows for preparing data for submission and using the Data Portal
  - **Network Data Management Plan**
    - Technical description of our data infrastructure
    - How the Data Portal is constructed, details of the standards implemented, security and data recovery procedures
    - Shouldn't be relevant to most users, but important to operations

Principles

Foster Open Science

- Timely and comprehensive exchange of ideas critical to success
- Data documentation and sharing are foundational

Accommodate Diverse Data Needs

- Variety of formats and use cases
- Sensitive and restricted data

Balance Benefits and Obligations

- Comply with grant and regulatory agency requirements
- Establish clear standards and expectations
- Enhance research visibility and impact
- Provide robust infrastructure and support

Terms

Research Data: Data collected and produced by ResNet scientists and ResNet funded projects

Data Publication: Process of preparing and releasing research data for use by others

Data Portal: Web-based platform for the publication and discovery of ResNet Research Data

Data Contributor: Individual conducting research and producing data.

Research Group: Network members affiliated with a Theme or Landscape.

Data Steward: Responsible for coordinating Data Contributors within a Research Group. (Co-)lead or delegated by lead of Landscape or Theme.

Data Administrator: Network Administrative Centre and Data Manager.

Roles and Responsibilities: Data Contributor

Conducts research and produces data

- Communicate expected data outputs from scientific research to Research Group Data Steward. Update as necessary.
- Document all data inputs. Create basic metadata where missing. Comply with existing usage requirements/policies for data inputs.
- Ensure accuracy of results to the best of ability.
- Declare and document data sensitivity or agreements with external parties.
- Package data and author comprehensive metadata in accordance with Network policies.
- Submit outputs to Data Portal for publication. Coordinate with Data Administrator as needed.

Roles and Responsibilities: Data Steward

Coordinates Data Contributors within a Research Group

- Foster the active participation of network members in the data publication process.
- Coordinate the efforts of Data Contributors within a Research Group.
- Report expected data outputs and progress of Data Creators within the Research Group to the Central Team on a regular basis.

Roles and Responsibilities: Data Administrator

- Establish data and metadata standards
- Review data submissions and prescribe revisions as needed
- Manage data infrastructure and moderate publication through online data portal
- Compile reports of data activities within Research Groups and communicate to the Network at large.
- Provide training and support for data publication
- Ensure the security and preservation of published data

Publication: Metadata

- Critical for data discovery and reuse
- "Data about data"
  - Description
  - Keywords from shared thesaurus
  - Sources and license information
  - Processing steps/methods
  - Attribute definitions
  - Persistent identifier (DOI)
  - Contact information
- Standard schema adopted by Administrator
  - Comply with regulations
  - Interoperability
- Administrator provides support, but Contributor responsible for authoring

Contributor Requirements

As early in project as feasible:

- Provide preliminary description of expected data outputs and revise as needed

Prior to project completion:

- Submit comprehensive, validated metadata

Metadata Training Friday, October 29 12-1pm EDT

Publication: Data

- Common data types (e.g. tabular, spatial)
  - Administrator defines standard formats
  - Stability, interoperability across research groups
  - Provide documentation, training, support
- Other types (e.g. transcripts, multimedia)
  - Contributor and Administrator work together to determine format

- Contributor packages completed data and submits to Data Portal
- Administrator approves or prescribes changes
- Publically accessible after embargo
  - Large datasets may need to be available on request
  - Some considerations may prohibit or limit access

Publication

Code

- Access to code created to process and analyze data is an essential component of open and reproducible research
- Should be preserved, documented, and packaged with research data whenever possible

Licensing

- Complex and evolving issue
- Strive for open-access
- Administrator will curate common licenses and assist
- Contributor must communicate
  - Restrictions inherited from data inputs
  - Sensitivity of the data produced
  - Agreements with external parties

Exceptions: Common Types

- Traditional Knowledge
  - Rights of traditional knowledge holders supersede the policy
  - Contributor responsible for engaging with holders to establish appropriate agreements
    - Existing frameworks such as OCAP should inform process
- Human Subjects
  - Privacy and confidentiality must be protected
  - Access in accordance with applicable legislation, regulations, ethics approvals and policies
- Sensitive Data
  - Where data release may cause harm, specific aspects of the data may need to be kept protected
    - e.g. locations of nests of endangered birds, sacred sites
- Existing Restrictions
  - Restrictions on preexisting data must be honored
- Other
  - Other exceptions will be reviewed by Administrator on case-by-case basis

Exceptions: In Practice

- Important for Contributors to identify and report exceptions
- Contributors should make best effort to participate
- Work together with Administrator for best solution
  - Some datasets may still be shareable
  - Fuzz/anonymize data
  - Password protected access
  - Publish metadata only

Bottom Line: You don't always have to share data, but we need to know it exists.

Usage: Embargo

From the Network Agreement:

- Data will belong to the Landscapes and Themes which generated them.
- Network members in these areas will have exclusive use for a maximum of six (6) months beginning from when the data is fully-acquired.
- Data will be communicated to the Administrative center as rapidly as possible.
- Network members will then have 12 months of privileged, password-protected access to the data.
- Afterwards, data will be shared with the public on the Network website to be managed by the Network data manager.

Usage: Acknowledgment

- Users of ResNet data must acknowledge data creators, contributors, and sources
- Should appear as formal citation
- Include Digital Object Identifier (DOI) when available

Arsenault, R., Brissette, F., Martel, JL. et al. A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds. Sci Data 7, 243 (2020). https://doi.org/10.1038/s41597-020-00583-2

Mitchell, Matthew G. E.; Bennett, Elena M.; Gonzalez, Andrew (2015), Data from: Forest fragments modulate the provision of multiple ecosystem services, Dryad, Dataset, https://doi.org/10.5061/dryad.41r51

Lu, Z., De Silva, A., Amarualik, P., Djibril, YS., Iqaluk, D., Jantunen, L. and Xie, H. (2020) Microplastics, perfluoroalkyl substances (PFAS) and organophosphate esters (OPE) pathways to Canadian Arctic lakes in Yukon and Nunavut via snow deposition. Waterloo, Canada: Canadian Cryospheric Information Network (CCIN). https://doi.org/10.21963/13195

Agriculture and Agri-Food Canada, 2020, "Annual Space-Based Crop Inventory for Canada, 2020", Agroclimate, Geomatics and Earth Observation Division, Science and Technology Branch. https://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9

Governance

Compliance

- Network funding is contingent on participation in data publication in adherence to this policy

Dispute Resolution

- Mediation and arbitration process established by Network Agreement

Amendments

- Modifications to policy must be approved by Science Advisory Committee
- Board will be informed of any changes

Feedback: Policy Components

What are the strengths, weaknesses, opportunities, and risks of this policy?

Are there any gaps? What mechanism/infrastructure needs to be implemented?

Up Next: Data Publication in Practice

https://bit.ly/ResNetDataFeedback

Data Publication in Practice

Data Publishing Pathways

- Data policy intended to be flexible
- Walk through examples of data publication
  - What are the constraints?
  - What is the process?
  - What is the outcome?
- Brainstorm other cases

Example: Open Access

Scenario

- Sam's project examines relationships between forest stand structure and biodiversity
- They spend a long field season establishing monitoring plots and taking inventories

Constraints

- Sam should benefit from novel dataset they've created
- Wants to collaborate with other researchers (and their data) through the data portal

Pathway

- Sam submits metadata and data to portal
- Authors manuscript during embargo
- Data publicly accessible after embargo

Outcome

- Sam publishes a great paper citing their own data
- Sam's data are findable, accessible, interoperable, and reusable by the public
- Sam's grant agency is happy
- Jo is thrilled to discover Sam's data and incorporates it in their large scale study, properly attributed

Example: Restricted Access

Scenario

- Alex is interested historical shifts in agricultural practices
- Alex works with First Nation to document oral histories

Constraints

- First Nation must retain Ownership Control Access and Possession of their data

Pathway

- Alex engages with knowledge holders to craft appropriate agreement
- Knowledge holders consent to data portal serving as their repository
- Administrator facilitates self-management through data portal

Outcome

- Metadata discoverable by all
- Only authorized users access data through Portal using assigned credentials
- Knowledge holders continue to manage access over time

Example: Metadata Access

Scenario

- Jamie's research examines how the creation of public open spaces influences the housing market
- Jamie reaches out to a large real estate company with extensive records
- They sign a data agreement providing access to a database of financial transactions

Constraints

- Commercial data is strictly proprietary, terms of their license prohibit the distribution of derived products

Pathway

- Jamie notifies the Data Admin about the restrictions
- Jamie includes detailed description of data attributes and methods in metadata
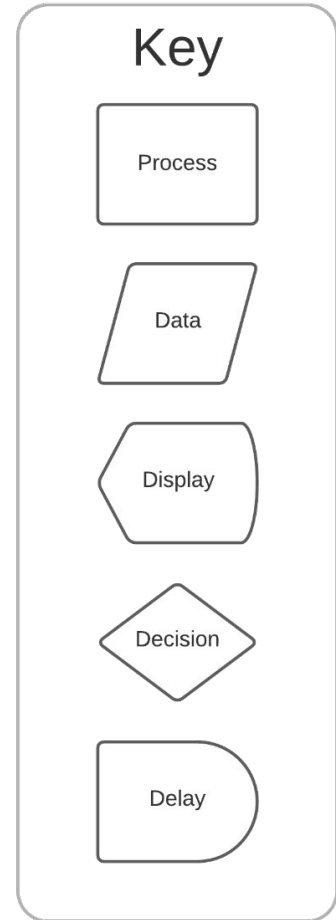- Submits only metadata to the portal

Outcome
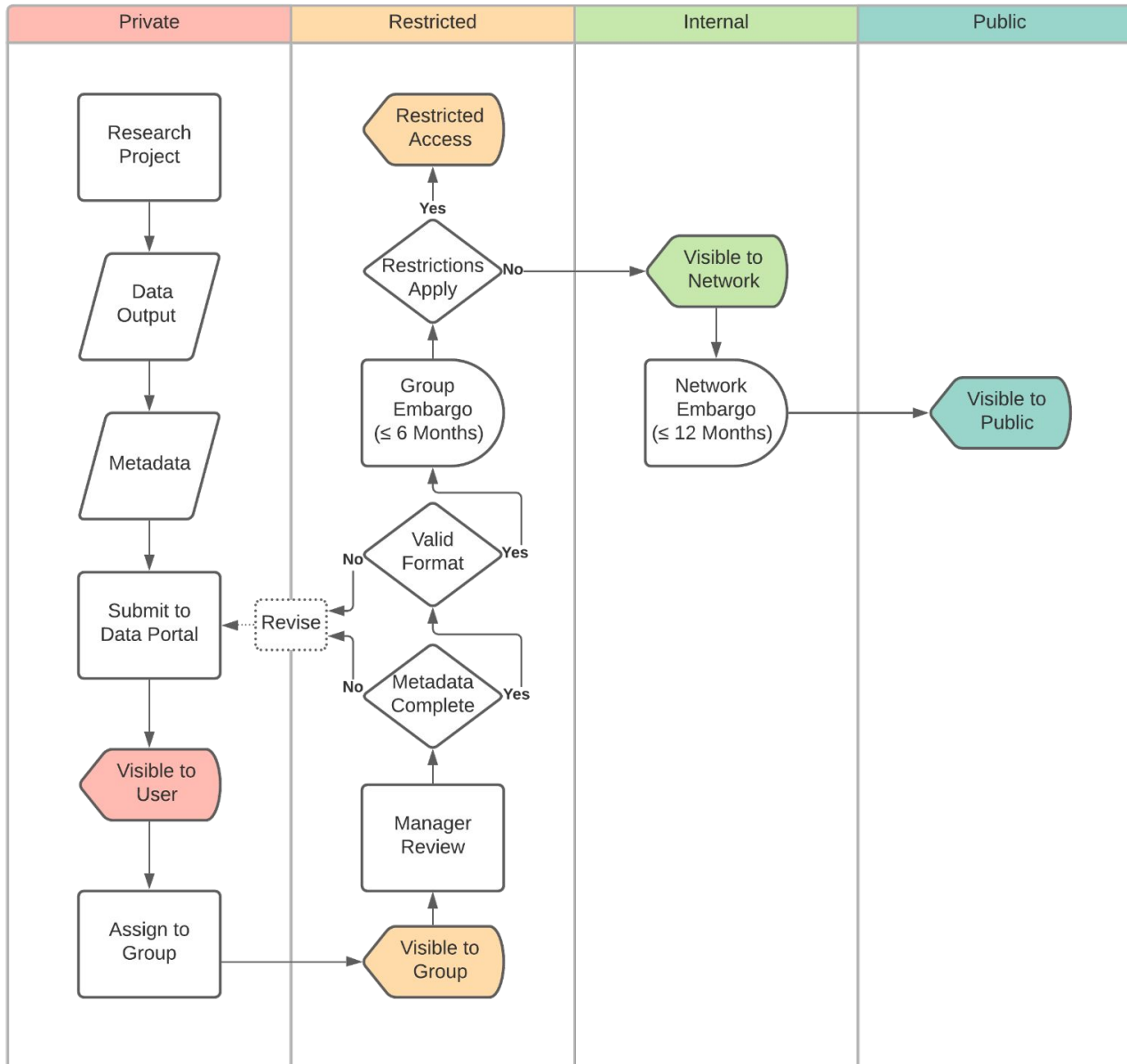
- Data not shared with network per existing agreement
- Metadata public
- Using Jamie's detailed metadata, Lauren is able to able to replicate the analysis in a different region using an alternate data source

Feedback: Publication Scenarios

- Brainstorm additional scenarios
  - What are the constraints?
  - What is the process?
  - What is the outcome?

https://bit.ly/ResNetDataFeedback

Data Publication Workflow and Visibility

Private | Restricted | Internal | Public

Private:
- Research Project
- Data Output
- Metadata
- Submit to Data Portal
- Visible to User
- Assign to Group

Restricted:
- Restricted Access
- Restrictions Apply (Yes → Restricted Access; No → Visible to Network)
- Group Embargo (≤ 6 Months)
- Valid Format (No → Revise; Yes → Group Embargo)
- Metadata Complete (No → Revise; Yes → Valid Format)
- Manager Review
- Visible to Group
- Revise

Internal:
- Visible to Network
- Network Embargo (≤ 12 Months) → Visible to Public

Public:
- Visible to Public

Key
- Process
- Data
- Display
- Decision
- Delay

What's Next?

Feedback

- https://bit.ly/MetadataPoll
- https://bit.ly/ResNetDataPolicy
- Engage on Slack (#data)
- Reach out directly

Workshops

- Metadata Training:
  Friday, October 29 12-1pm EDT
- Data Portal Orientation:
  Thursday, November 18 1-2pm EST

Data Policy

- Incorporate comments and feedback
- Bring before Science Advisory Council
- Send out to network if approved

No need to wait until policy finalized, get in touch now!

Data Office Hours

Everything data: discovery, processing, analysis, visualization, development, publication, puns…

All welcome!

Wednesdays 2-3pm ET

"Knock" on my door via Slack/Email and we'll set up a Zoom session

Contact

John Clark
NSERC ResNet
Data Manager, Dashboard Developer
john.clark3@affiliate.mcgill.ca