

PERMAFROSTNET DATA POLICY

Version 1.0 – 2020-04-01

Edited and compiled by Nick Brown nick.brown@carleton.ca

1 INTRODUCTION

NSERC PermafrostNet is a strategic partnership network that connects knowledge holders across complementary domains to understand the interactions of climate, permafrost and infrastructure. The network will produce knowledge products to support improved adaptation to climate change.

This is Version 1 of the PermafrostNet data policy that is expected to evolve over time. This policy is not legally binding and is intended to complement and provide further specificity to section 8 of the NSERC PermafrostNet Research Agreement.

2 VISION

Our vision for permafrost data in Canada is that it is openly accessible in a way that supports the needs of stakeholders and respects the rights of communities. Canada has dedicated cyberinfrastructure that supports researchers and practitioners to curate and share their data.

PermafrostNet will lead the development of prototype systems toward this vision and will:

- Make permafrost information and data more accessible by facilitating exchange among permafrost researchers and user groups, including northern communities and international programs.
- Support the adoption of data standards and the creation of an interoperable system of permafrost data providers.
- Encourage a cultural change towards best practices in research data management

3 SCOPE AND TERMS

In this policy, the term *data* includes those data collected and generated by the PermafrostNet community when those data have benefited from access to network resources. This includes data collected in the field, in a laboratory or through simulation. It also includes data that have been homogenized, cleaned, or processed by members of the PermafrostNet community to be more accessible.

This document also references *open data*. The most important details of what it means for data to be open are summarized in the open data handbook [1]

The term *code* includes source code developed with the benefit of network resources.

This policy also refers to the PermafrostNet community. For the purposes of this document, the PermafrostNet community includes co-applicants, collaborators and any HQP funded by the network.

4 PRINCIPLES

The PermafrostNet data policy is based on the following set of guiding principles:

- Respect the rights of data and code creators and knowledge holders.

- Support full, free and open access to data and code with minimal delay so that all sectors (communities, academia, industry, government) may benefit and contribute to gaining further insight.
- Ensure long-term preservation of data and code.
- Recognise that additional restrictions may apply to some datasets including data involving traditional knowledge or containing sensitive information.
- Use existing knowledge and infrastructure where appropriate and continue to develop an interoperable, connected data system.

5 ACCESS

5.1 Metadata

Discovery-level metadata¹ which describe a dataset shall be made open and accessible on the shortest possible timescale. Data may be made available using network digital infrastructure when applicable or using existing catalogues². When selecting a catalogue, we encourage data creators to use those that provide DOIs such as Polar Data Catalogue (polardata.ca) or Nordicana-D (cen.ulaval.ca/nordicanad/).

Network funding may be withheld if discovery-level metadata are not made publicly accessible.

Use metadata¹ which describe the content of datasets must be made available along with the data (5.2).

5.2 Data

Data shall be made open and accessible immediately after the expiration of an embargo period. The standard embargo period for data shall be 12 months following the collection of the data and may be extended.

Extensions to the embargo period may be requested from granted by the Scientific Committee and must take into consideration the network principles of equity, diversity and inclusion. The purpose of extensions is to support researchers to publish results from data they collect or process. As such, requests for extensions must include a clear description of the research questions that the data are intended to support and how the delay of data release supports answering those questions.

Data may be made available using network digital infrastructure when applicable or using existing data repositories².

Data shall be made available to members of the PermafrostNet community within the shortest possible time to support collaboration within the network, under the provisions outlined in section 6.2.

Highly structured data including geotechnical measurements and time series of temperature or subsidence, should be made available in a format that can be integrated into the Permafrost Information Network of Ground Observations, adhering to the most recent PermafrostNet standards for

¹ Discovery-level metadata provide searchable information that enables data consumers to find datasets online, for example using search engines or libraries. Discovery-level metadata describe the who, what, where and when of the dataset as well as conditions for access, use and reuse. *Discovery-level* metadata contrasts *use* metadata which allow a data user to understand the data. [5], [6]

² Examples of such repositories include the Polar Data Catalogue and Nordicana-D.

permafrost data, which are to be incrementally published on the PermafrostNet website (permafrostnet.ca).

6 USAGE OF DATA

6.1 Acknowledgement

To recognize the valuable role of data providers (and scientists who collect or prepare data) and to facilitate repeatability of experiments in keeping with the scientific method, users of PermafrostNet data must formally acknowledge data authors (contributors) and sources. Where possible, this acknowledgment should take the form of a formal citation or reference, such as when citing a book or journal article.

If not specified, citation should include as many of the following as possible:

- DOI(s)
- Author(s)
- Publication date
- Title
- Editor or compiler
- Publication place
- Publisher
- Distributor
- Distribution location
- Access date
- Data within a larger work

When DOIs are provided together with data, they must be included in the citation or reference.

6.2 Before open release

Because data are made available to network participants before open release, and because the publication of derivative works using these data may jeopardize publications by the data creator(s), data users are required to notify data creators and receive their permission before the creation of any derivative works during the embargo period. This notification should take the form of a brief description of how the data are to be used. In the case where the data creator has students who may be affected by the release of the data, the data creator is expected to consult with those students before granting permission.

Because data may not be citeable before open release, data users producing derivative works from data before its open release must also ensure data creators are properly recognized for their contributions. The following guidelines are adapted from the ArcticNet data policy, which was modified from the CASES Data Management Plan and the Canadian JGOFS formula:

A. Early drafts of any publication (e.g., presentations at meetings, papers, book chapters) must be made available to each data creator responsible for one or several of the data sets used in the publication. The data creator then consults his/her research associates, post-doctoral fellows, students, and technicians to assess eventual rights to authorship.

B. Within three weeks of reception of the draft, data creators involved in the paper must inform the senior author of their wish to contribute (or not) to the publication, to see themselves or members of their team included in the list of co-authors, or to be mentioned in the Acknowledgements.

C. In cases where data creators and the senior author or original co-authors of the draft cannot reach an agreement, the conflict will be resolved by the Scientific Committee.

As well, and in all fairness, co-authorship should not be automatically justified for technical or clerical contributions to a central dataset repeatedly used in a large number of studies, or when the dataset represents only a small fraction of the derivative work, particularly when the removal of the single dataset would not have an appreciable impact on the derivative work.

6.3 Allowable data use

Users of data provided by PermafrostNet may use the data in any medium, mode or format for any lawful, commercial or non-commercial purpose subject to the terms of this data policy.

6.4 Licensing

Open data should be released under an appropriate licence which, at a minimum, are conformant with the principles set out in the open definition. PermafrostNet currently recommends a Creative Commons public domain or attribution licence³, or the Open Data Commons equivalent. This recognizes that legal enforcement of attribution is not always possible, and that providing a citation and asking for credit is consistent with academic norms.

6.5 Dispute mediation

Resolutions to any disputes shall be proposed by the Scientific Committee. Should the disputing parties not be able to agree to the proposed solution, the Board of Directors will have the final decision, weighing the recommendation of the Scientific Committee and a possible rebuttal.

7 EXCEPTIONS TO DATA POLICY⁴

7.1 Traditional Knowledge

Where local and traditional knowledge is concerned, rights of the knowledge holders shall not be compromised and shall supersede data policy regulations presented here.

It is the responsibility of researchers to engage respectfully with applicable traditional knowledge holders on a project-by-project basis to establish appropriate agreements regarding the collection, usage and dissemination of data. Existing frameworks and principles including OCAP[2] and NISR[3] should be used in the development of these agreements. Agreements made with traditional knowledge holders shall supersede data policy regulations presented here

³ Data licensing is complex, and this recommendation is expected to change over time. Additional licenses may be added to datasets. For more information on the recommendation, see openaire.eu/research-data-how-to-license/. For a list of other possible licenses, see opendefinition.org/licenses/.

⁴ These exceptions were adopted from the Canadian IPY 2007-2008 data policy

7.2 Human subjects

Where human subjects are involved, privacy and confidentiality must be protected. Access to personal information and human biological samples may be provided in accordance with applicable legislation, regulations, ethics approvals and policies including the relevant Tri-Council Policy [4].

7.3 Sensitive data

Where data release may cause harm, specific aspects of the data may need to be kept protected (for example, locations of nests of endangered birds or locations of sacred sites).

7.4 Existing restrictions

Where pre-existing data are subject to access restrictions, those restrictions must be honoured.

7.5 Other exceptions

All other exceptions are to be reviewed by the Scientific Committee.

8 CODE

In the interest of promoting open and reproducible science, and in advancing the technical capabilities of the broader research community, code that is developed with the benefit of network resources must also be made open and accessible so long as this does not infringe on intellectual property rights (section 11). The schedule and requirements for making code open and available should follow the guidelines set out for data in sections 6 and 7.

Code should be released under an appropriate licence which, at a minimum, allows for reuse. PermafrostNet currently recommends a GPL licence⁵, recognizing that more permissive licences such as MIT can be added if needed. Licences should not restrict use by network partners in industry.

As a best practice, any code that is released should include the capabilities necessary to recreate the environment and dependencies required to run it. This can be done through containerization using software such as Singularity or Docker, or by including a list of the packages necessary to run the code.

9 NETWORK RESPONSIBILITIES

9.1 Retention and security

PermafrostNet will endeavor to secure and back up data that is stored by the network to mitigate the risk of data loss, within the limits of its capacity. Opportunities and needs for sustainable data management or archiving of large collections will be investigated no later than the beginning of Year 4 of network funding (2022-2023).

9.2 Provision of service

PermafrostNet will prototype services to support the vision outlined above. This may include the development of customized tools to process legacy datasets or the creation and hosting of databases for network data. Services are provided on a best-effort-basis and cannot be guaranteed.

⁵ Licensing is complex, and this recommendation is expected to change over time. For a list of other possible licenses, see <https://spdx.org/licenses/>.

10 RESPONSIBILITIES OF DATA CREATORS

1. When making data available, raw or minimally processed data must be included. For example, in the case of temperature measurements, providing yearly averages exclusively would not be acceptable if data were collected hourly. However, cleaning up a raw instrument output file to make it more machine-readable would be acceptable if metadata are not lost in the process.
2. Data creators are responsible for ensuring that the data they make available are, to the best of their knowledge, free from errors.
3. Data creators must follow network data standard when applicable. Where the network data standard does not provide guidance, data creators are expected to use their best judgement.
4. Data creators must acknowledge PermafrostNet when publishing data created with the benefit of network resources.
5. DOIs should be created when making data available to facilitate citation, version control and the automatic notification when data are used or referenced.
6. In the spirit of open data, proprietary formats are strongly discouraged when making data available.

11 INTELLECTUAL PROPERTY

With respect to intellectual property, this policy shall defer to the PermafrostNet Network agreement, the relevant parts of which are copied in Appendix I

12 DISCLAIMER

Data provided by the network are distributed “as is”, and both the network and the data creator exclude all representations, warranties, obligations, and liabilities, whether express or implied, to the maximum extent permitted by law.

Neither the network or the data creator is liable for any errors or omissions in the information, and will not under any circumstances be liable for any direct, indirect, special, incidental, consequential, or other loss, injury or damage caused by its use or otherwise arising in connection with this licence or the data, even if specifically advised of the possibility of such loss, injury or damage.

13 REFERENCES

- [1] Open Knowledge Foundation, "Open Data Handbook." [Online]. Available: <http://opendatahandbook.org/>. [Accessed: 26-Nov-2019].
- [2] First Nations Information Governance Centre, *OCAP Path To First Nations Information Governance*. 2014.
- [3] ITK, "National Inuit Strategy on Research," 2018.
- [4] Canadian Institutes of Health Research Natural Sciences and Engineering Research Council of Canada and Social Sciences and Humanities Research Council, "Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans," 2018.
- [5] Earth Science Information Partners (ESIP), "Attribute Convention for Data Discovery 1-3," 2017. [Online]. Available: http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery. [Accessed: 23-Jan-2020].
- [6] NorDataNet, "Types of Metadata." [Online]. Available: <https://www.nordatanet.no/en/node/71>. [Accessed: 23-Jan-2020].

14 APPENDICES

APPENDIX I. INTELLECTUAL PROPERTY

The following is copied from the NSERC SNG PermafrostNet Research Agreement. Where this copy and the original differ, the

- All Arising Intellectual Property shall be governed by the institutional/organizational policies of the creator of such Intellectual Property.
- With the exception of copyrights, the Parties agree to grant to each other, a non-transferable, non-exclusive, royalty-free, fully paid-up, world-wide perpetual licence to use, reproduce, modify, adapt, or implement any Arising Intellectual Property for non-commercial purposes only. No rights to sub-licence are granted and nothing in this Agreement shall be construed as granting a licence to any existing Intellectual Property or rights. The Parties further agree to indemnify and hold all other Parties harmless of any costs, suits, claims, or damages resulting from the use or misuse of the licence.
- For clarity, authors shall maintain exclusive ownership of any and all rights, including moral rights, associated with any copyrighted works which they may produce in the performance of the project.
- The Parties agree that the authority and responsibility for making decisions with regard to legal protection and commercialization of Arising Intellectual Property shall rest with the owners of the Arising Intellectual Property. Where there are two or more owners of the Arising Intellectual Property, they shall designate an agent to act on their behalf. For greater certainty and without limitation, unless otherwise agreed to in writing on a case-by-case basis by the owners of Arising Intellectual Property, no one shall have any rights in the Arising Intellectual Property, other than the right to a non-exclusive licence provided for in clause 9.2 above.
- Pursuant to the mandate of the Strategic Partnership Grants for Networks Program, every reasonable effort must be made to have the results of Network Research exploited in Canada for the benefit of Canadians. Accordingly, the Parties shall act in accordance with the NSERC Policy on Intellectual Property.

APPENDIX II. NOTES ON THE DEVELOPMENT OF DATA POLICY:

Version 0.1

The development of this policy has been informed, in part, from a plenary session of the 2019 PermafrostNet AGM. Participants were first asked to think about what principles should guide the PermafrostNet data policy. They discussed this at individual tables of approximately five people each. Each table was then assigned to one of eight topics and asked to share their opinions at their table and identify areas of agreement or disagreement. Several Over the course of the data session at the 2019 AGM Plenary, there were a few items that were brought up multiple times:

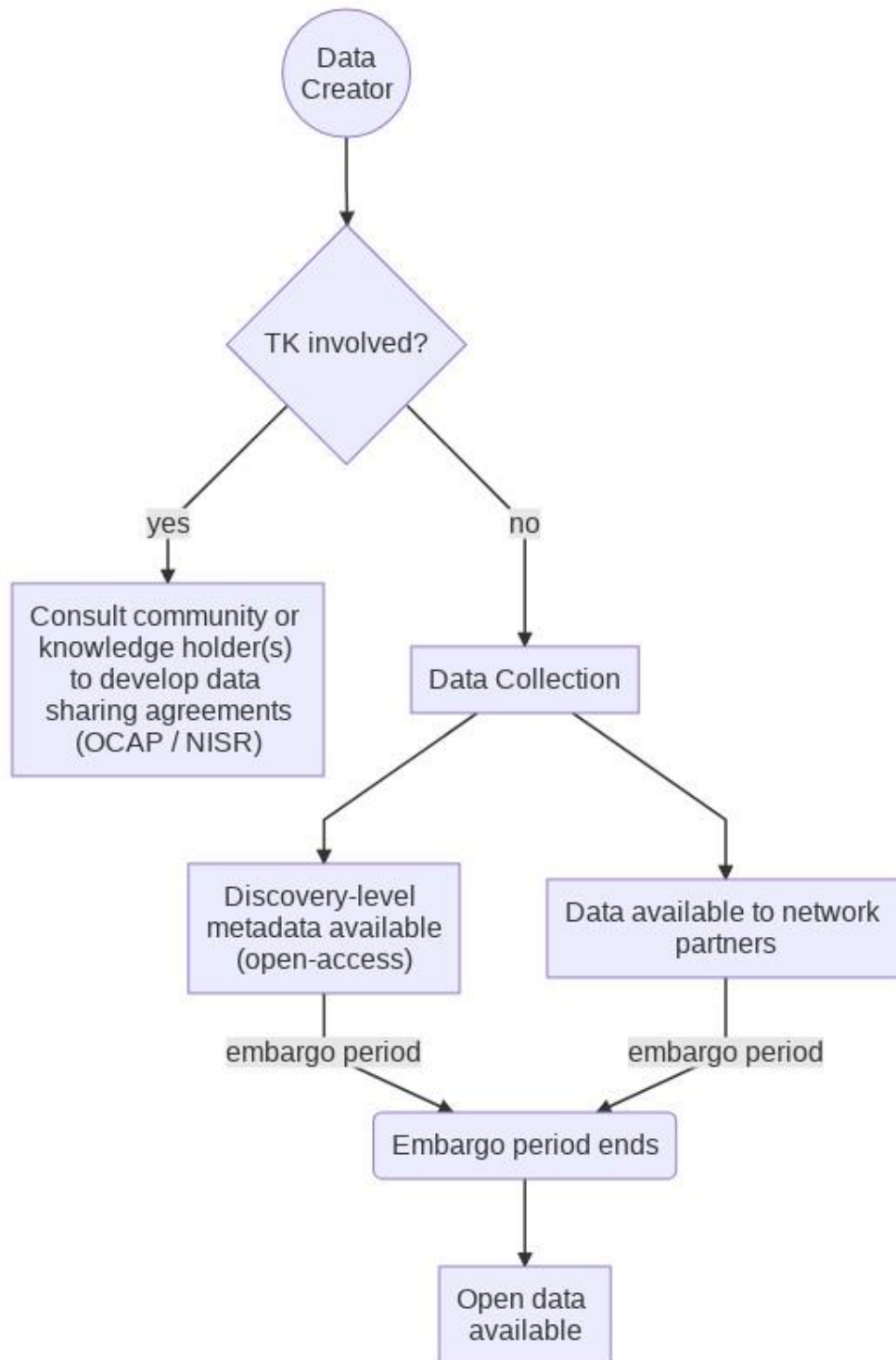
1. *“<Existing policy X> works well” and “don’t reinvent the wheel”*. Much of this data policy is based on the ArcticNet / Polar Knowledge Canada policies.
2. *“Traditional Knowledge must be considered”*. On the recommendation of the participants, this policy makes explicit reference to existing frameworks and principles including NISR and OCAP. When it comes to traditional knowledge, the default
3. *“Policy must be flexible to fit with partner organization or university policies”*.

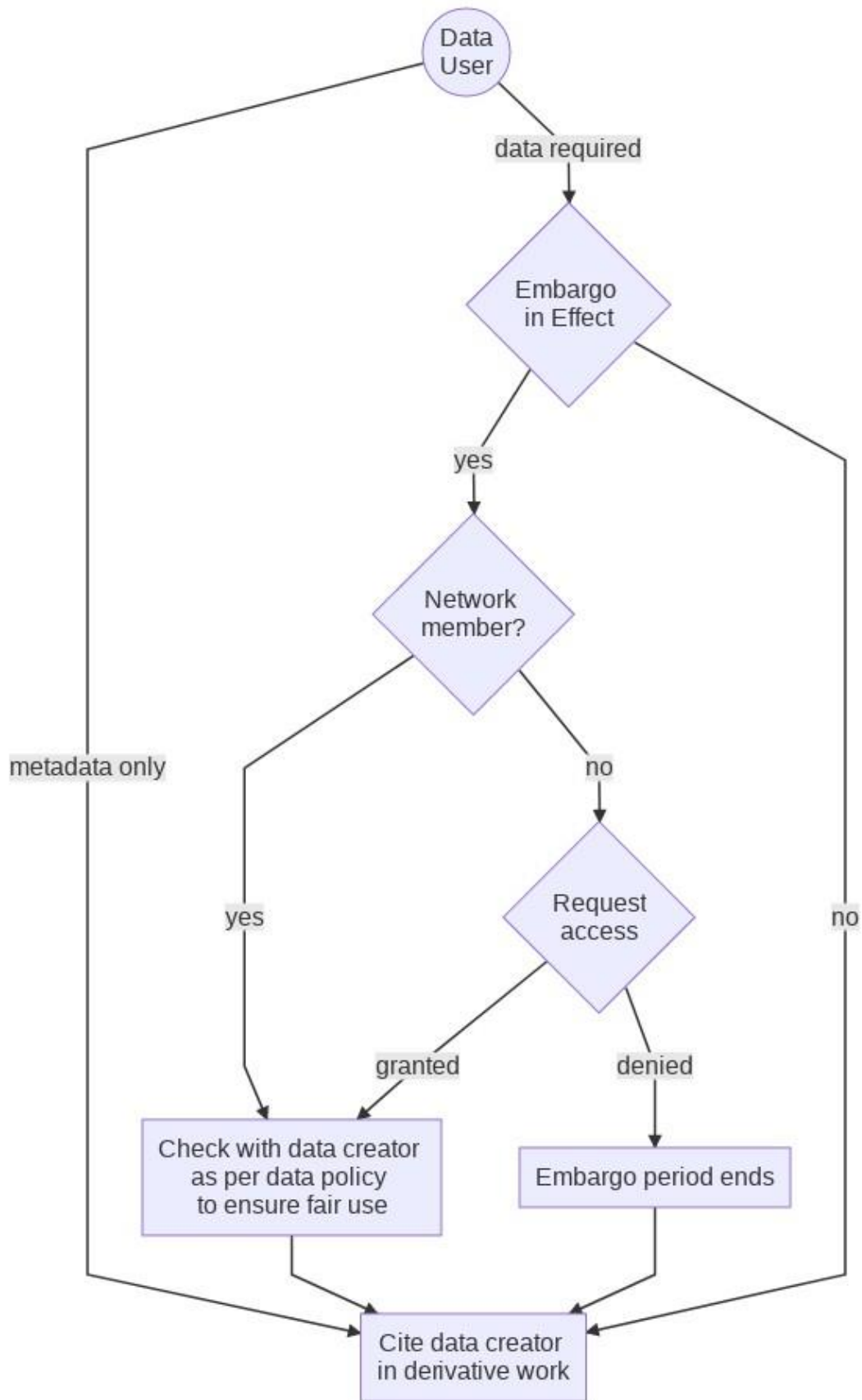
Finally, there was some feedback regarding what the responsibilities of data users should be. Several participants recommended that data be made freely open and available. Several participants suggested that data creators should be notified when data are used for derivative works to initiate collaboration. When considering existing policies, the use of CEN datasets does require notification, but many other policies (NSIDC, PDC, IPY) do not. GTN-P recommends inviting data creators to collaborate when using “regulated” datasets, but only recommends notification for open datasets.

This policy requires that data authors be notified during an embargo period prior to the free and open release of datasets. During this period, data is made available within PermafrostNet but not necessarily to the public.

Considering that one of the network outputs is a federated database system, accessible through the permafrost data science platform, it may be possible to automate author notification for data accessed through this platform. However, developing such a tool would represent a considerable investment of effort and would need approval of the SC.

APPENDIX III. FIGURES





Responses from AGM plenary workshop

Included or addressed in policy

Partially included or addressed in policy

Conflict or not included

Out-of-scope or not a question

<p>Openness, transparency, accessibility, reproducibility</p> <p>Minimum embargo period, other than making sure the data is up to standards</p> <p>Principles of reproducibility for data</p>	P
<p>Open access to data to provide knowledge to public</p> <p>Ensure accessibility in official languages as well as local language</p> <p>Access to computer code for reproducibility, innovation</p>	P
<p>Policy</p> <p>Open access</p> <p>Partners use: NDA, level of access – timing</p>	B
<p>Policy</p> <p>Respect all parties rights & interest</p> <p>Quality control / QA</p> <p>Data & code separation</p>	B
<p>Policy</p> <p>IP / Patent</p> <p>Acknowledge</p> <p>NSERC policy</p> <p>First nations OCAP</p> <p>QA/QC on Data</p>	B
<p>Responsibilities of data creators</p> <p>Universal template of data and metadata QAQC of the data when possible, build on existing conventions</p> <p>Be clear on uncertainty and potential data biases, context and methods of data collection</p> <p>Follow principles of reproducibility, existing or established by the network</p> <p>Acknowledge the network if the data was collected as part of the network, making it available in a timely manner</p>	P
<p>Data policy-</p> <p>Policy to address data producers, users and those who disseminate info</p> <p>What are the data QAQC standards and who is responsible for this</p> <p>Liability / Disclaimers?</p> <p>What about archived data where metadata standards aren't met?</p>	G
<p>PermafrostNetData – Policy should be flexible but tending towards open access</p> <p>Data should be made available to network participants: ASAP</p> <p>Metadata should be public: ASAP</p> <p>Allow long embargos but ultimately require open publication</p> <p>TEK data policy should follow local protocols and legal documents</p>	G

Update the arcticnet statement on TEK to reference existing indigenous protocol(s)	
<p>Data</p> <p>Open access</p> <p>Are publications to be open access only?</p> <p>Data to be submitted at end of projects (sub project of network)</p> <p>Submit to Nick</p> <p>Nick provides format for submission of data (temperature / geotechnical)</p> <p>Scientific committee can authorize hold back of data before publication</p>	G
<p>Acknowledgements</p> <p>Use CEN data policy : Acknowledge the dataset / author (cite as paper)</p> <p>Notify author, initiate collaboration</p>	G
<p>Scope</p> <p>Need to be flexible re: HQP</p> <p>Funded HQP and other HQP who may be working on related projects (but not funded via network)</p> <p>Partner organizations contributing already established data sets – may have their own policies</p> <p>Learn best practices from other research network (one recent work with eee)</p>	R
Is there a public government-supported database for open-access?	B
Data policy needs to be flexible to accommodate different partner policies (gov't, indigenous, university)	B
TK – needs to be self-determined @ project level taking into consideration principles like OCAP (first nations) NISR (national Inuit strategy for research)	B
<p>Data policy</p> <p>Time for open release? Data embargo?</p> <p>How do you slice what's been driven by permafrostNet funds vs not? ← (scope)</p>	B
<p>Data policy principles</p> <p>Data standard should be considerate of NRCAN, PIN, territorial datasets etc</p> <p>Scope is key question to determine things like:</p> <ul style="list-style-type: none"> • Open • Timely – once QAQC is done and data are cleaned up (within 2 years) <p>TK – use storage, compilation & publishing of TK up to the Inuit /FN community</p> <p>Self determination by particular FN/ indigenous communities / organizations / governments on a project level</p> <p>Interoperability of web portals / databases</p> <p>CCADI OCAP principles ITK principles, Inuit knowledge entre</p> <p>ITK research strategy</p>	B
<p>What data is proprietary and what is public?</p> <p>How do we get consensus</p> <p>Protocols on public statements</p>	B
<p>Data policy ideas</p> <p>How open and timely will it be?</p> <p>Metadata</p> <p>Focus on really good metadata, mandatory</p> <p>Arcticnet works well</p> <ul style="list-style-type: none"> • Produce metadata: essential • 1st step: open source right away, can link to other data source, allow to contact author if data not yet posted • Metadata is most important for the database 	G

Metadata Let people (or give people the option to) submit to the repository that already works for their data type Make submission of metadata mandatory (arcticnet: no metadata submission = no money)	G
Open access data With embargo on availability of data itself to give the person who collected the data the chance to publish it (avoid: student goes on maternity leave, returns to find that A.W.I. has published global data work that includes local data) Also, notify data author and initiate collaboration	G
Metadata Types of data made publicly available: "raw", "data summaries", "publications" Quality control methods for data Data sharing agreements and collaboration agreements Network vs government data submission / storage? Field data vs data logging vs computer simulation?	G
Indigenous Knowledge There is an ever-expanding body of knowledge on indigenous knowledge and data and information. Familiarizing permafrostnet with this work over time would be important CCADI has a working group led by ITK focused on this topic	B
Usage Data should be public Room for companies to develop products with the data Need access to view data, available to wider audience Uptake is the end goal Usage, intricately tied to acknowledgement and standards / data quality, traditional knowledge, legal requirements	P
Open access First nations OCAP principles of open access	Y
Retention: Short-term: CU Need a hand-over strategy or policy Long-term: NRCAN? GSC? ECCC?	Y
Acknowledgement Co-authorship if field data is used How is the contributions of field data? For data generated for PermafrostNet there is a distinction between members of the network (co-authorship) vs external use.	Y
Recommend using existing archives Good experience with arcticnet system Software with dependencies and code changes through time necessitates consideration Since DOIs are available, that makes data sharing easier	-
Ensure data is available to various users in different ways: scientists, universities, schools in the communities, children. Strong link with mobilization Different ways to present data and log who is using what data and for what. For example, a school in Nunavut downloads specific data set. Tiered approach	-

Help desk – can help users navigate what data is required Increase impacts	
1. Open by default: follow IASC statement on principles 2. Attribution: publication, DOI etc.	-

notes