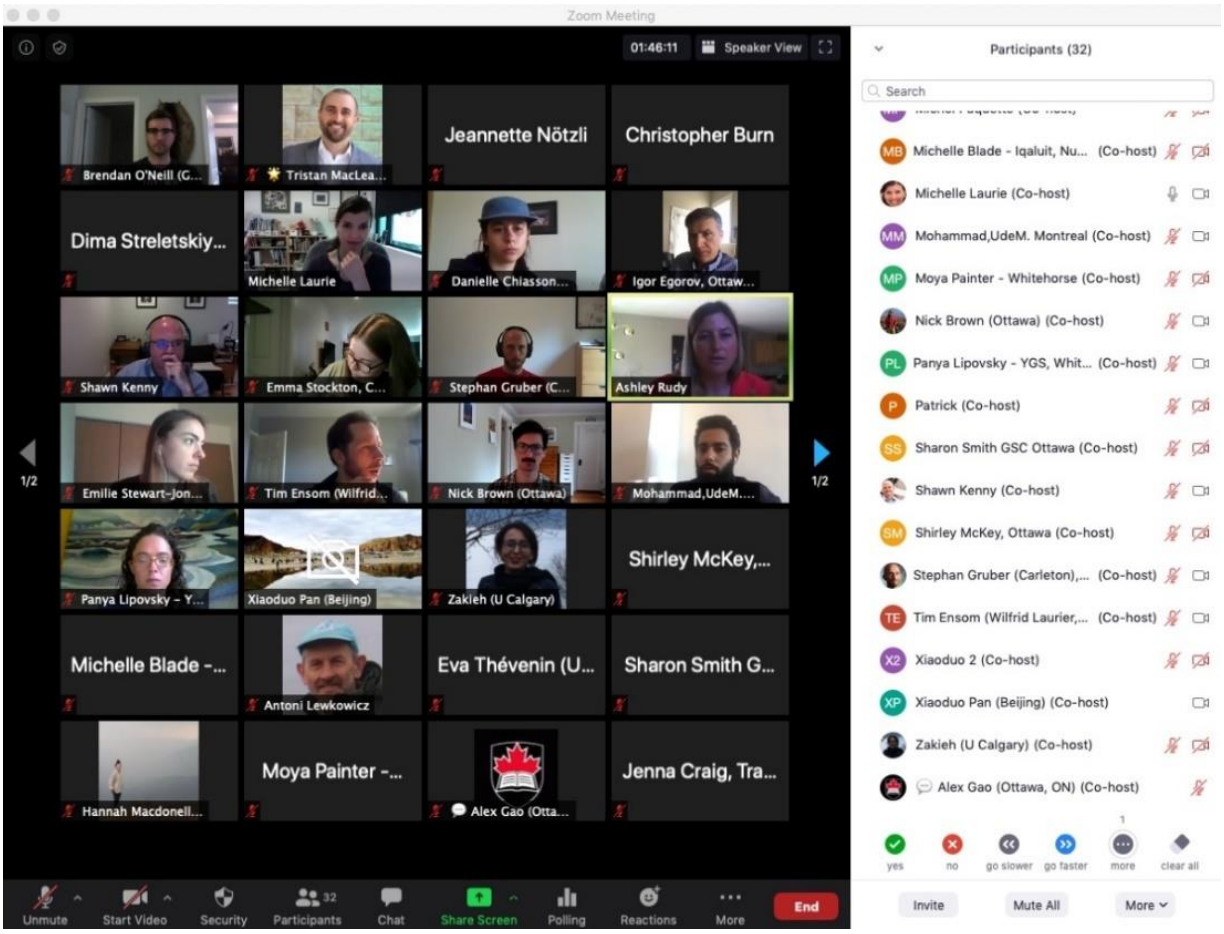
Permafrost Data Workshop 2020: Final Report

NSERC PermafrostNet



Citing this report:

Brown, N., Gruber, S., Pulsifer, P., Stewart-Jones, E. (2020) Permafrost Data Workshop Final Report. NSERC PermafrostNet. May 27-29, 2020: Ottawa, Canada. DOI: 10.22215/pn/10120001 .

Table of Contents

1) Summary and background:

The issue of permafrost data management, accessibility and harmonization has been a longstanding concern within the permafrost community. Previous permafrost workshops in 2013[1], 2017[2] and 2018[3] have highlighted the need to make permafrost data more accessible, and the need for support at all levels of the data life cycle including: quality control, access, publishing, referencing, interoperability and archiving. This workshop was designed to build on previous workshops by adding specificity and context to the data needs that have been established.

The main themes that emerged include:

- **Capacity-building is needed at all levels**: participants at all levels of permafrost data management highlighted that there are insufficient resources available for the significant task of gathering, cleaning, and disseminating permafrost data.
- **Using existing tools and solutions** is important to conserve resources, avoid the duplication of effort and ensure long-term sustainability.
- **Access to standardized data** and **discoverability of data** by remains a challenge**.**
- **Continued communication** will be essential to ensure that data systems are not developed in isolation.

2) Workshop participation

The workshop took place May 27-29, 2020 digitally via the Zoom communications platform. A total of 81 participants joined from five provinces, three territories and five different countries (Canada, USA, Switzerland, China and Japan). The event was originally planned to be an in-person meeting held in Ottawa but was moved to an online-only format, due travel restrictions during the COVID-19 pandemic. To compensate for this, the workshop made use of several digital collaboration platforms to recreate the in-person workshop experience as much as possible, including: Zoom, for videoconferencing; Mural, to display digital flipcharts and sticky-notes; ThoughtExchange, for collecting and ranking survey information; and Google Docs for collaboratively writing content. The workshop had strong participation across stakeholder groups including students, coinvestigators and network partners (Figure 1, Appendix III).

[1] *Pan-Territorial Permafrost Workshop* https://www.northernadaptation.ca/sites/default/files/pan-territorial_permafrost_workshop_report_0.pdf

[2] *Towards a Canadian Permafrost Network*: https://carleton.ca/permafrost/wp-content/uploads/WorkshopPermafrostNetworkOttawa2017_-Report.pdf

[3] *Pan-Northern Meeting on Permafrost Hazard Mapping* https://northernadaptation.ca/sites/default/files/summary_report_final_compressed.pdf

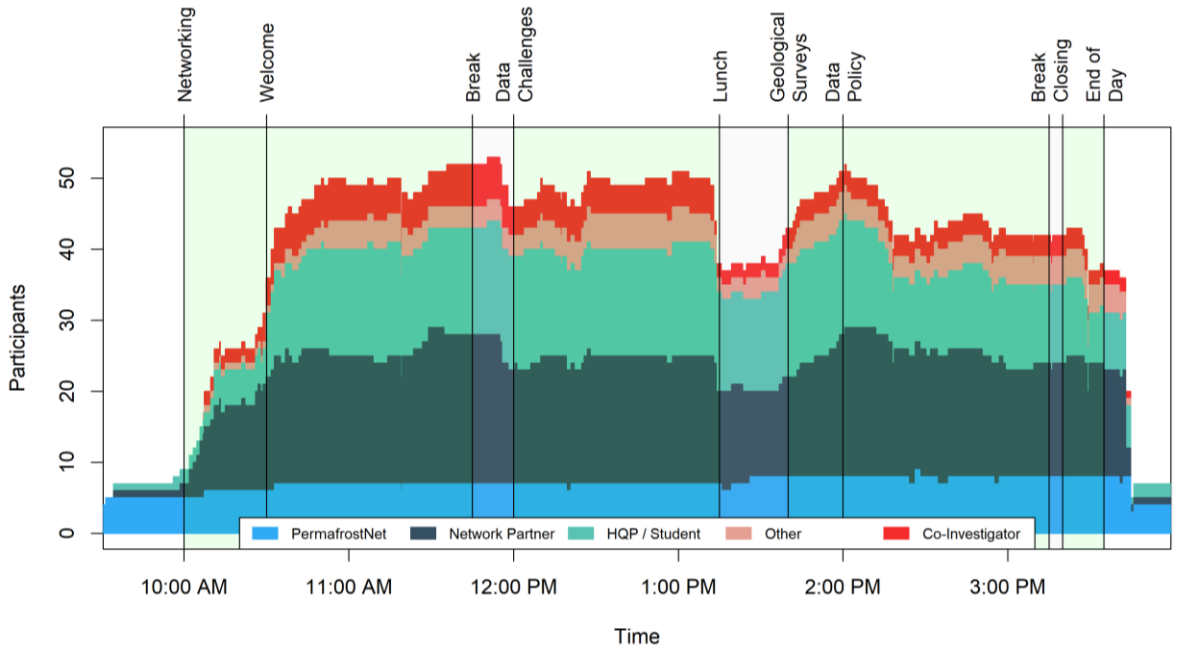


Figure 1: Participation on day 1 of the workshop (May 27, 2020) as determined by Zoom records and grouped by affiliation to PermafrostNet. Participation charts for all three days are shown in Appendix III.

The workshop featured presentations from national and territorial geological surveys (Geological Survey of Canada, NWT Geological Survey, Yukon Geological Survey) about their respective permafrost data and data systems. There were also presentations from individuals representing international permafrost or cryospheric data centres, including the Swiss Permafrost Monitoring Network (PERMOS), the Tibetan Plateau Data Centre (TPDC), and the Global Terrestrial Network for Permafrost (GTN-P). A presentation on semantics and the FAIR data principles provided participants with background knowledge to contribute on day 2. Copies of these presentations are available on the attendee's area of the workshop website[4]

Overall, the workshop demonstrated strong northern involvement and commitment to permafrost data management (Figure 2, Appendix III). The workshop represented a diversity of views, approaches, and levels of prior knowledge. Although a multi-day workshop requires substantial effort, this broad engagement was effective in creating "buy-in" and support to move forward with a unified approach to permafrost data management at a national level. It also helped to create a shared vision and to elevate the technical and conceptual understanding of the broader permafrost community.

[4] https://www.permafrostnet.ca/resources/events/permafrost-data-workshop-attendees-area/

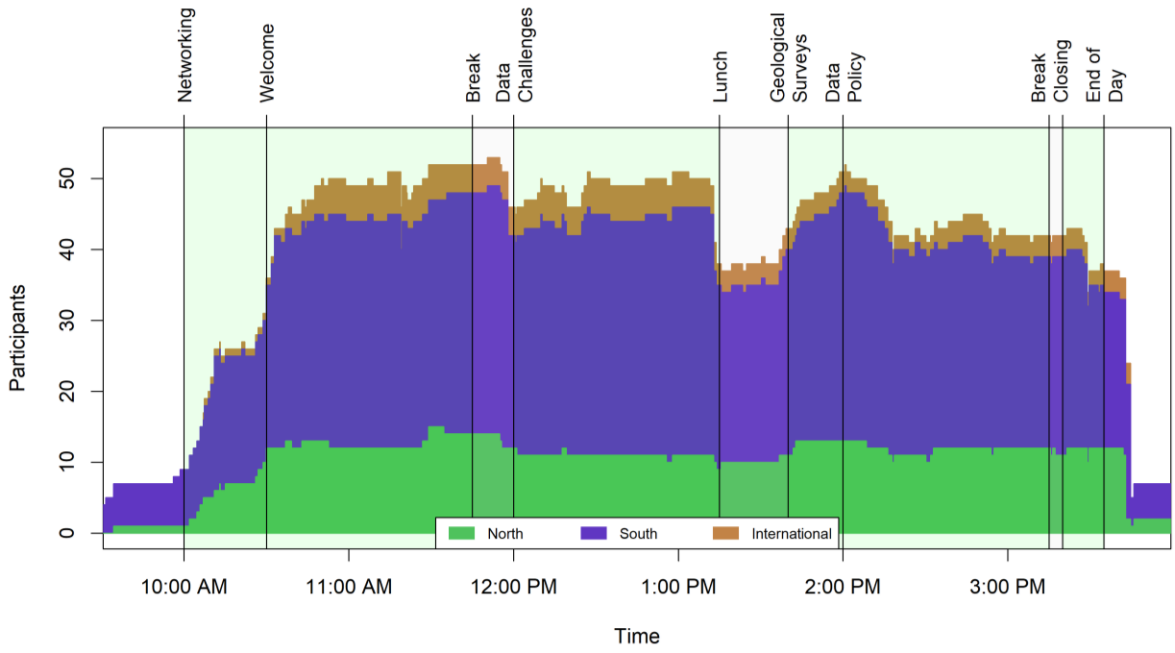


Figure 2 Participation on day 1 of the workshop (May 27, 2020) as determined by Zoom records and grouped by participant location.

3) Acknowledgements

This event would not have been possible without the help of the following people and organizations:

Michelle Laurie provided strategic process design and facilitation for the sessions. Tristan MacLean managed the technical aspects of running a virtual meeting to make sure things ran smoothly. Emilie Stewart-Jones, Tim Ensom, and Emma Stockton were rapporteurs for the sessions, taking notes and summarizing key ideas. Danielle Chiasson, Alex Gao, and Hannah Macdonnell helped moderate the chat.

Thanks also to everyone who volunteered their time to present or participate in a panel: Alex Bevington, Ashley Rudy, Boyan Brodaric, Dmitry Streletskiy, Étienne Godin, Jeanette Noetzli, Joe Melton, Matt Jones, Nick Brown, Panya Lipovsky, Peter Pulsifer, Sarah Brown, Stephan Gruber, and Xiaoduo Pan.

The workshop was put on with the help of the Canadian Consortium for Arctic Data Interoperability (CCADI) and Transport Canada. Finally, nothing would have been possible without the participation and enthusiasm of everyone who attended. A complete list of participants is included in Appendix II.

4) Identifying needs, challenges, and solutions:

Pre-workshop survey

In advance of the workshop, participants were invited to participate in a survey, which asked: *Imagine its May 2021. What happened at this data workshop that was really important to you? What did we achieve or produce?* Participants could then rank ideas generated by others. A total of 41 people responded to this question, generating 27 individual thoughts with 312 total ratings. These were themed into categories, the most popular of which was making permafrost data more interoperable.

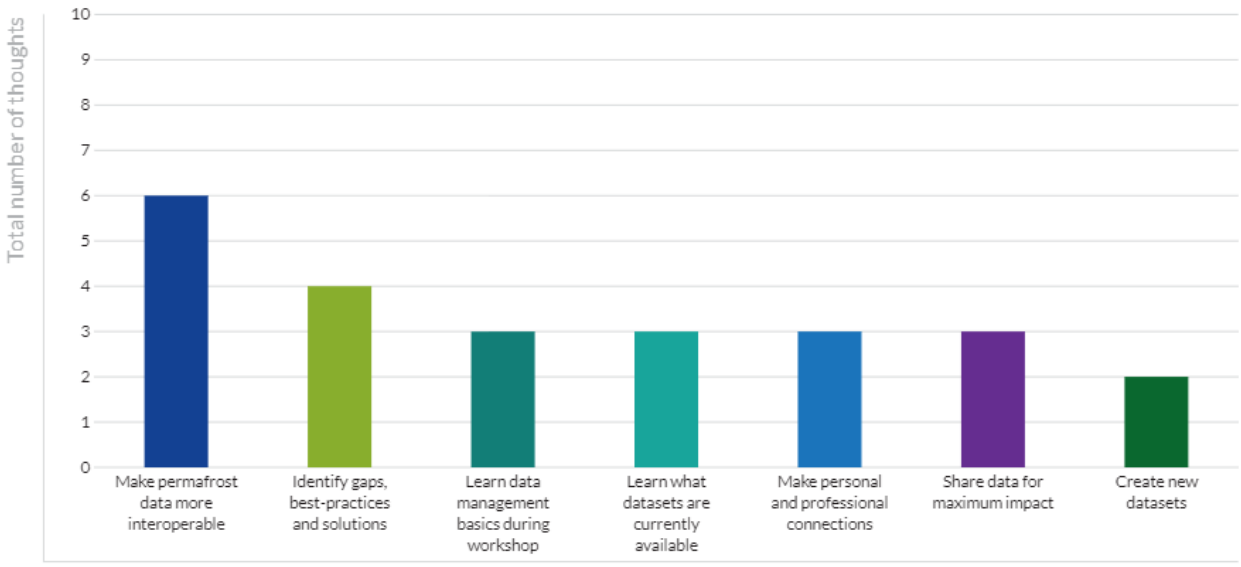


Figure 3: Results from pre-workshop survey grouped by theme.

Welcome session

The goal of the first morning session was to set the scene for the Permafrost Data Workshop and create a shared understanding of the workshop's purpose by discussion of shared challenges and opportunities. The morning began with a panel discussion to answer the question: "*In your experience what's the biggest challenge you face in accessing, using or sharing permafrost data"*

Table 1: Panellists for the introductory session

Ashley Rudy	NWT Geological Survey
Sarah Brown	NWT Association of Communities
Joe Melton	Environment and Climate Change Canada
Alex Bevington	BC Ministry of Forests, Lands and Natural Resource Operations and Rural Development

Throughout the discussion of challenges to accessing, using, and sharing permafrost data, a number of recurring themes emerged:

1. **Data standardization**. How should we go about creating data standards and what are the important considerations? How should I (as a data collector) format my data for others? How do we harmonize data from multiple sources?
2. **Limited resources**. Cleaning and formatting data require a great deal of time, which is often in short supply both for researchers and even organizations responsible for data management
3. **Data sharing**. How do we raise awareness and get buy-in across the community? How to create incentive for sharing data?
4. **Sustainability**. How do we sustain a permafrost database or data repository in the long run?

Some specific challenges unique to different user groups were also brought up (Table 2)

Table 2: Specific challenges raised during the introductory session

Who	Challenge or need
Government	Standardizing data (from multiple sources) so that it is useable requires time, and resources are limited for this task.
Government	It is difficult to find permafrost data within government.
Modeler	I need access to large amounts of data from a single standardized source.
Modeler	Variability in permafrost metadata makes it hard to search. We need standardized metadata.
Communities	Community members and workers do not know where to find permafrost data.
Field scientist and data user	Data (metadata) changes over time, so we need a way of accounting for environmental change over time (rocks falling from rock wall where a sensor is installed).
Scientist	It is important to preserve original data (raw data and metadata).
All	Different levels of data processing are required for different user groups.

Challenges and solutions for permafrost data

The next session was designed to identify shared problems within the permafrost community, leverage network connections to help identify what is needed to help with the data creation process, and to identify use cases for data. After organizing into small groups, participants posted their challenges, solutions and ideas using different coloured sticky-notes on a digital flipchart (Figure 4, Appendix IV). This session provided a forum for participants to share experience and demonstrated the value of operating as part of a network.

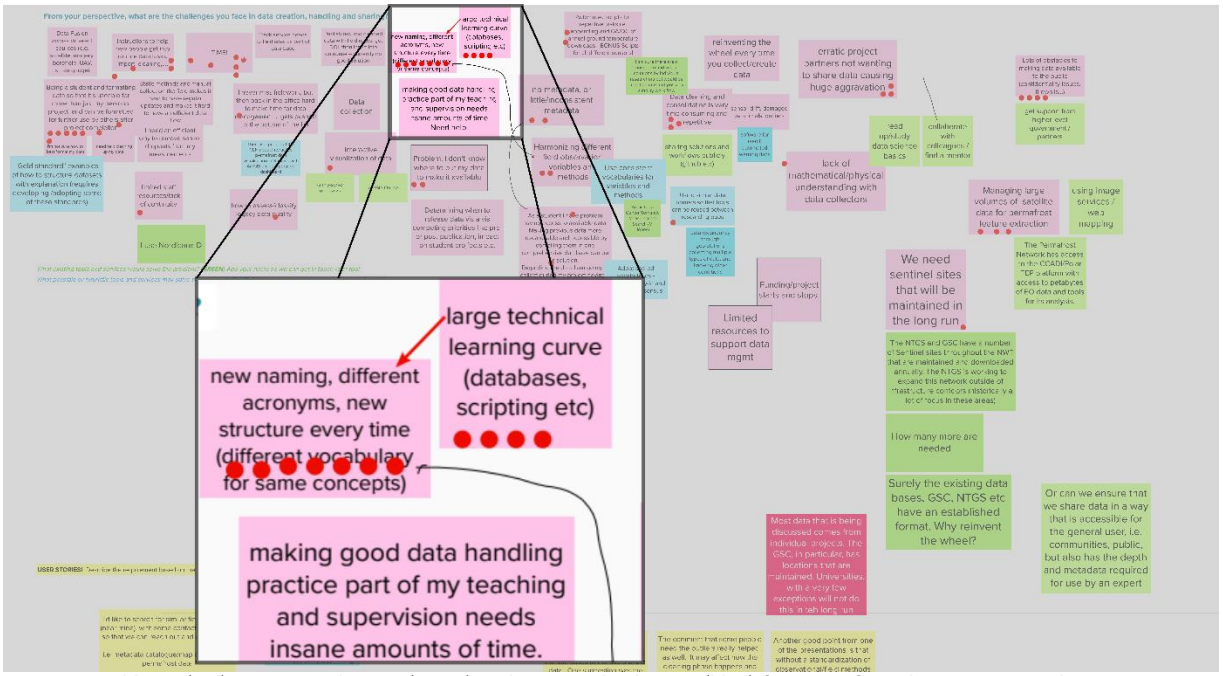


Figure 4: Problems (red), existing solutions (green) and proposed solutions (blue) for permafrost data creation and curation. Participants added red 'stickers' to indicate priority items. Zoomed detail-view shown for clarity. A full-resolution version is available in Appendix IV.

Throughout the discussion (verbal and written) on the problems and challenges with data creation and access, *a number of* recurring themes surfaced:

- **Limited resources**: both time and funding
- Data **inconsistencies**
- **Need for instruction** on data management
- A number of general **barriers** to data sharing

Highest among participant's challenges was finding the time for data management. For governments, the trouble manifests as a lack of continuity in data management projects due to inconsistent funding. A proposed solution to this problem was **sharing data cleaning and formatting tools and workflows** such as automated scripts. **Developing universal software** for these tasks would be the longer-term goal. In both cases, making the data more homogenous and/or interoperable will be an important first step so that the automation of certain workflows is feasible.

Inconsistencies in vocabularies, data collection methods, data structure, and metadata remain major obstacles to data interoperability. Two solutions were suggested: (1) **Adopting shared vocabularies** with community buy-in and scientific consensus and (2) **developing standards and guidelines** for cleaning, formatting, and structuring data within the network.

Although solution (2) would provide guidance to network students collecting and formatting their data, the large technical learning curve to data management tasks such as scripting and working with databases still requires a solution.

Finally, problems around data sharing were wide-ranging, from questions about how and where to share data to concerns around releasing data. Solutions to the former issue were to **create a separate data repository** for the network or to **work with existing repositories** to provide storage for datasets and/or to generate DOIs**.** Because the permafrost community may have needs that are not currently met by any existing repositories, (such as access to data as a OGC WMS service), a hybrid approach may be desirable in which some data are kept on PermafrostNet servers. If any data are held exclusively by PermafrostNet, they will likely need to be transferred to a more stable location at the conclusion of network funding.

Table 3: Specific scenarios and needs

Who	Scenario or need
Field researcher	"I'd like to search for similar field locations (near mine), with some contact information so that we can reach out and collaborate. i.e. metadata catalogue/map of Canadian permafrost data."
Decision makers, consultants, and less technical users	"Provide higher level summary data […]"
Data interface developer	"I need to understand the use cases of the users of the data"
Researcher	"I need 'raw' data for certain studies"
Field researcher	"Track service needs to field sites as part of database"
Field researcher	"I need an efficient way to remove sensor dropouts from my measurements"
Field researcher	"Provide a location to host our raw data quickly at the end of the field season, then we can take our time to clean/update"

Field researcher	I need tools to deal with sensor drift, sensor damage and outliers in temperature time series data

A short discussion on how these challenges can be met in the short- and long-term elicited responses from two participants, however no specific consensus was achieved.

Timeframe	Participant 1	Participant 2
Short term	Developing a web map of available permafrost data with metadata	• Developing processing tools for the primary data types: ground temperature and geotechnical data • Making the other data types available without standardization
Long term	Creating a universal set of tools for automation of data processing	

5) Network data policy

This session focused on two elements of the NSERC PermafrostNet data policy that were flagged during community consultation in winter 2020: (1) the adoption of a data quality management plan to ensure the network is a trusted source for data and (2) the issues of open data release. Both topics also have implications beyond the network for the long-term needs to support permafrost data management in Canada.

Data quality

The first part of the session dealt with discussions of what is needed for data quality management within the network and beyond. This could have implications for professionals who use network data. In other data policies considered, data quality was either not mentioned (GTN-P, NSIDC), explicitly not guaranteed (Nordicana-D, GWF, ArcticNet) delegated to the researchers submitting the data (GWF, ArcticNet), or quality checked by the organization accepting the data (PERMOS).

Although the session began with a focus on establishing data quality criteria, the general agreement by the end was that making well documented data available can be more valuable than establishing a formal quality control system. In short, data “quality” is context-specific and is difficult to define for all user groups.

Participants were asked to brainstorm individually and in small groups, a ‘dream list’ of all the elements in an ideal data quality management plan for permafrost data. They returned to plenary and took part in a survey to identify the minimum specifications for this plan by posting items from their dream list and rating the ideas that others suggested.

Table 4: Examples of data quality elements grouped into themes

Theme	Participant suggestions for a data quality management plan
Metadata: General	• Metadata (location, provenance, uncertainty, etc) is required for datasets and data elements (e.g. observations) • Inclusion of metadata • Clear Metadata Standards

Metadata: Data quality levels or lineage	• After data processing has been done, outline what steps have taken place • Include data provenance and methodology • Rank data quality (letter grade) • "Levels" of data cleaning (level 0 = raw, level 1= cleaned)
Metadata: Site Conditions	• Dream: site conditions (vegetation, snow, water, ...) • Qualitative comments: Notes in the data on site changes over time.
Software or Visualization Tools	• Automated filters: Thresholds, outliers, etc. • Visualization: Easy visualization of data will help find problems in the data • Tool(s) or methods to detect gross or “out of range” errors
Raw Data	• Raw data: Keep raw data in database • Raw data is good so that processing can be repeated • Imperfect data can be useful
Long-term considerations	• Consistent and simple system that can be maintained over the long term • Long-Term management: legacy use archiving and use

The most common dream list suggestions were thematically related to a need for metadata, software or visualizations to help perform quality control, and the preservation of raw data. In addition to the general need for metadata, two specific categories of metadata emerged as common recommendations: descriptions of site conditions, and a description of the processing steps or processing levels. Tools for data checking were more popular in the dream list than in the final list of minimum specifications.

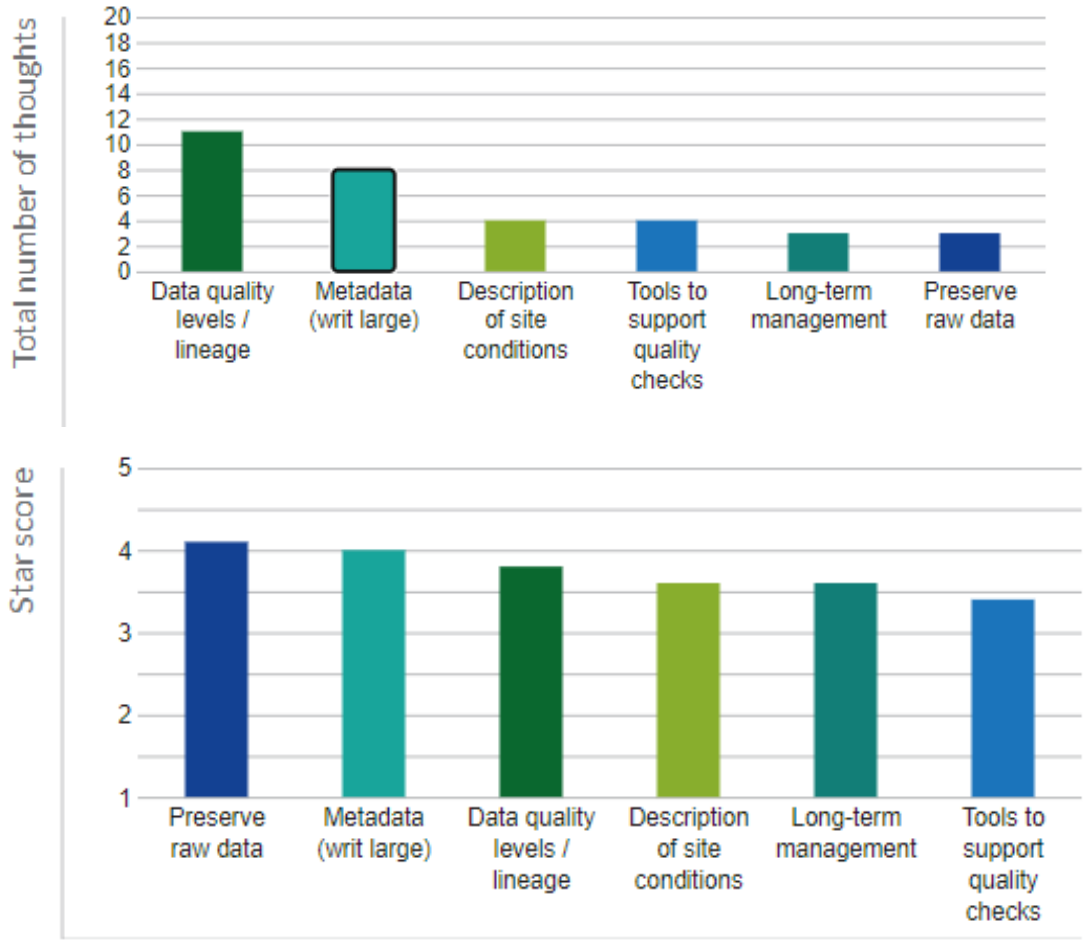


Figure 5: Total number of thoughts in each theme (top) and ranking of data quality elements as essential chosen by workshop participants (bottom).

Other thoughts that did not fall into a clear theme included having a feedback mechanism for users to report issues with the data, and including a disclaimer with the data to reduce any liability. Some of the un-themed thoughts were more closely related to broader principles of data interoperability than quality *per se*, such as the "*use of FAIR principles",* having *"open data, not private data"* and requiring *"Lots of documentation to make sure that data quality management is consistent".*

The results of this session suggest that core elements of a data quality management plan require having the **appropriate metadata** to judge whether data are fit for a specific purpose and having **access to the raw data**. Software tools are important to support quality control, and are desired by the community, but are less essential for ensuring that network data can be trusted.

Open release

One of the aims of releasing data openly is to make it useful to people outside of the network. To explore this issue from another perspective, a technique called TRIZ was used to flip the question around; participants were asked to identify all the possible ways to achieve the most undesirable outcome, namely: *"What can we do to ensure that data generated by the network is not findable, accessible or reusable by people outside the network?"*. The TRIZ method is thought to allow groups to challenge conventional practices and discuss topics that are difficult to discuss.

Participants created a list of possible behaviours that would lead to this undesirable outcome, then added stars next to things that they do or that they have experienced in any way, shape, or form. The most commonly reported behaviours fell under the themes of intentionally restricting access, publishing data late and systemic policy issues.

Table 5: Behaviours that the community can undertake to achieve the undesirable outcome of making sure data generated by the network is not findable, accessible or reusable by people outside the network. Higher numbers of stars are related to behaviours that are more commonly experienced or impede the reused of data most strongly.

Theme	Participant Examples	Stars
Intentionally restrict access	• Keep data offline • Password-protected • Unwilling to share data	**24**
Publish data late	• Take years to release data • Not sharing data until after end of research project, PhD, MSc, …	**23**
Systemic issues	• Require multiple levels of approvals to release data • Have 1 central gatekeeper	**20**
Inappropriate format	• Share data in a proprietary format that requires paid software • Use Excel formatting to store data (i.e. blue = data from sensor 1)	**14**
Don't allocate time or resources to data	• Ignore data requests • No manager of legacy datasets • Plan to go on vacation immediately after fieldwork	**13**
Inadequate metadata	• Raw data with calibration missing • No metadata	**13**
Legal restrictions	• Require a signed agreement to ensure liability • Require confidentiality clauses	**8**
Don't release raw data	• Only share processed data	**5**
Keep data offline	• Locally sharing data (USB drive, hard drives passing by hand)	**5**
Don't follow standards or conventions	• Create new language • Invent your own standard because finding out about what exists is more work	**3**
Charge for data	• Charge huge amounts for each data point	**2**
Make data access difficult or unclear	• Do not publish metadata in a widely used data catalogue • Frequent changes to host location (i.e. server) or web front layout	**2**

Finally, participants were asked what the community can do – or stop doing – to achieve our desired objective of making network data useful to stakeholders outside the network. The solutions or

behavioural changes that were suggested included: **using technological solutions**, to keep data online or in shared locations, or store it in a decentralized way to increase access and limit data loss. **Developing carrots and sticks** such as introducing incentives for people to make data open, or penalties for not doing so. This was suggested both for NSERC PermafrostNet and for tendered contracts. **Making changes to institutional policies**, replacing requirements of data quality with descriptions of levels of quality, or using moderation-based rather than approval-based approaches to releasing data. Finally, **developing better communication** would ensure that best-practices are developed with a broad group.

6) Permafrost data interoperability

Priority data types for interoperability

This workshop was focused on ground temperature data and basic geotechnical properties (such as ground ice content, and grain size) so that the scope was narrow enough to make early progress. These data types were chosen because they are currently the most widely available and are foundational to permafrost research and monitoring. However, many other data types are relevant to the permafrost community.

Participants were invited to share their priorities for which data types should be the next focus for interoperability and standardization. This was done using a ThoughtExchange survey and grouped into themes. Like other surveys in the workshop, the difference in rank between the most- and least-popular choices was relatively low (2.9 – 4.1).

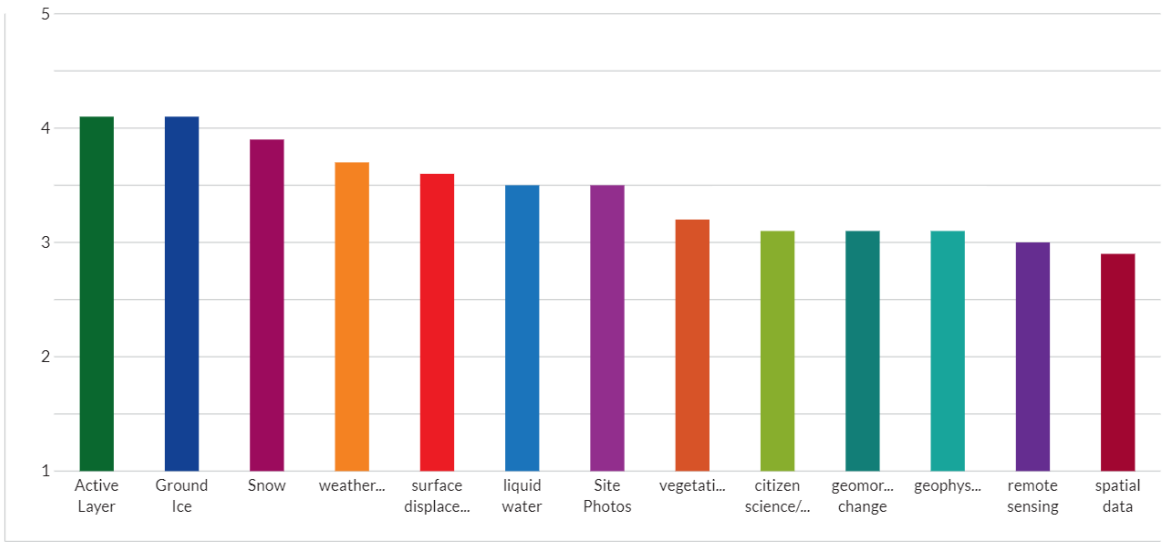


Figure 6: Participant priorities for data types to be the next focus for interoperability efforts.

This session generated a lively discussion about the best approach for identifying the next steps that should be taken, and the importance of distinguishing between established data collection or reporting conventions and data management strategies.

The polling / voting approach that was taken during the workshop was seen as important to gather broader perspectives and generate buy-in from stakeholders, but concerns were raised that a large group discussion is not appropriate for dealing with some of the technicalities. For people unfamiliar

with using a certain kind of data, such a discussion could lead to people tuning out of an online meeting. Suggestions were made to identify a motivated group of users and experts to make a proposal for others to comment on at a future event, and to prepare it in advance with a high degree of guidance.

In future discussions of interoperability and standards, the distinction must be made between those standards governing how data are generated or collected, and the standards or specifications describing how those data are shared, or otherwise made interoperable. The development of former, which already exist in some cases (such as describing ground ice in samples), are best left to technical experts in the respective scientific domains.

Vocabularies and semantics for permafrost

The inconsistencies in the vocabulary and semantics within the permafrost community impede data sharing. Establishing a shared vocabulary and adopting it as a community is critical to data interoperability. The purpose of this session was to (1) prioritize concepts for which the network and broader permafrost community want to adopt or establish controlled vocabularies or semantics, (2) describe the process and requirements for developing this shared language, and (3) develop a shared understanding of the available resources.

The session centered around discussion among an expert panel (Table 6) who discussed the question: "*What advice would you give the network to work effectively in adopting or adapting vocabularies and semantics for sharing permafrost data?*". After hearing the panelist discussion, participants were given time to identify concepts for which they thought shared vocabularies would be valuable in the short-term, then this list was given further consideration and comment from the panelists.

Table 6: Expert panellists for discussion on vocabularies and semantics

Boyan Brodaric	Natural Resources Canada
Matt Jones	Arctic Data Center, National Center for Ecological Analysis and Synthesis
Peter Pulsifer	Geomatics and Cartographic Research Centre, Canadian Consortium for Arctic Data Interoperability
Etienne Godin	Université Laval, Nordicana-D

The main recommendations from the experts were:

- Do not reinvent the wheel. Recognise previous and simultaneous efforts and engage with the broader community.
- Determine the degree of sophistication of the product (Figure 7).
- Start early to reduce effort later on.
- Split the terms from their definitions and focus on the definitions or concepts. This increases vocabulary flexibility and allows it to be multilingual.
- Permafrost experts must work together with technical experts.
- To promote engagement in the process and to reach agreement, the work on shared vocabularies needs to be done in the open.

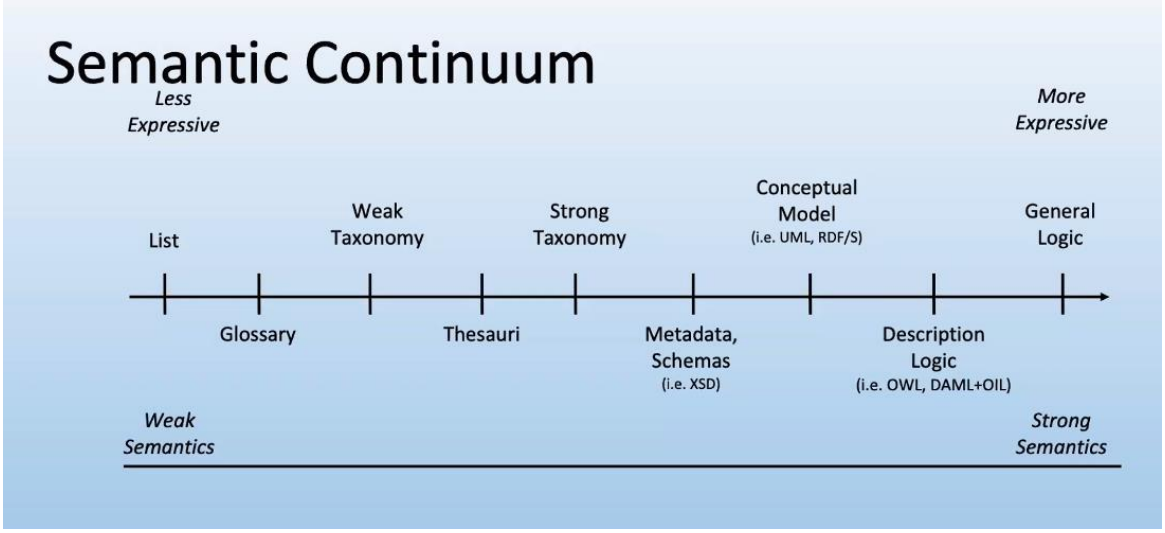


Figure 7: The semantic continuum, where the complexity and expressiveness increase from left to right (image credit: Peter Pulsifer, 2008[5])

Based on the discussion with the experts, the process for developing a shared vocabulary within the network or the broader permafrost research community must be iterative and involve permafrost domain experts as well as knowledge modelers and technical experts (Figure 8). There appeared to be agreement around using the Glossary of Permafrost and Related Ground-Ice Terms[6] as a starting point. It was suggested that an association such as the Canadian Permafrost Association (CPA) should lead this process and that the network be used to crowd source commentary and ideas or to provide committee members. To engage with the broader international community, it was proposed we have representatives engaging in the larger process (such as the ENVO ontology).

The development of a complete permafrost ontology would represent a significant investment in effort. An agreed-upon glossary may be a more valuable and more realistic interim goal for the Canadian permafrost community as it works to make existing data more available and interoperable. However, to be efficient with resources, these efforts should be aware of existing initiatives. Several of these were mentioned in the workshop as possible collaborators that could be valuable to engage with:

- **The Environment Ontology (EnvO)**, an international ontology designed for environmental entities such as ecosystems, environmental processes, and environmental qualities
- **The Ecosystem Ontology (ECSO)**, a DataONE ontology originally designed for Carbon Flux measurements for MsTMIP and LTER Use Cases
- **The World Wide Web Consortium** (W3C) suite of ontologies

[5] Pulsifer, P. L. (2008). *An Ontological Exploration of Antarctic Environmental Governance: Towards a Model for Geographic Information Mediation*. Doctoral dissertation, Carleton University.
[6] http://globalcryospherewatch.org/reference/glossary_docs/permafrost_and_ground_terms_canada.pdf

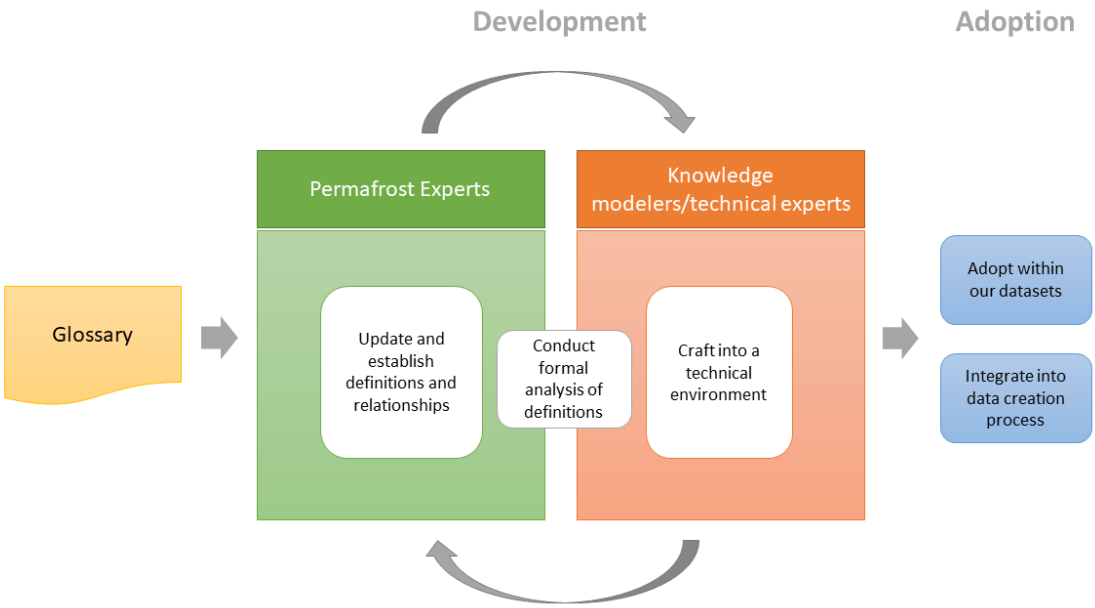


Figure 8: Process for developing a shared vocabulary within the permafrost community

Two workshop participants are engaged in efforts to improve arctic data interoperability and highlighted the importance of collaboration with existing international groups, namely:

- The Canadian Consortium for Arctic Data Interoperability (CCADI)[7]
- Arctic Data Committee (ADC), Vocabulary and Semantics Working Group[8]
- The Arctic Data Center[9]

Standards and specifications

Adopting shared standards and specifications will make data sharing more efficient. In this session, participants generated a list of conventions, standards, or specifications for permafrost data that they had used or encountered (Table 7). This session focused mostly on CSV-formatted data because of its ubiquity in the permafrost community.

Table 7: Existing conventions or standards for permafrost data brainstormed by participants.

	Standards	**Comments**
Time	Unix time, Geologic time, Julian Day, Date and time in two separate fields, YYYY-MM-DD HH:MM:SS, ISO 8601	Radiocarbon and relative dating do not fit into these standards
Time zone	UTC, Include timezone, ex: UTC-7	

[7] https://ccadi.ca/
[8] https://arcticdc.org/activities/core-projects/vocabularies-and-semantics-wg
[9] https://arcticdata.io/

CSV structure	Long-format (Entity-Attribute-Value), Wide-format, RFC 4180	Format changes from data collection, to processing, to storage, to download
Depth	Meters, Meters with decimals	Is depth of borehole vertical or normal to surface?
Coordinates	UTM coordinate system, Latitude and longitude in DMS vs DD	
Missing values	Negatives (when it's not temperature) NA, NaN, Fixed values (6999, 9999, -9999)	
Units	Carbon content: Kg C / m Temperature: Degrees Celsius International System of Units (SI)	
Variable names	CF Standard Names	

In a follow-up discussion, participants evaluated the merits of adopting one or more of the identified standards as a community. There were:

Time: The ISO 8601 is an unambiguous and broadly used datetime standard that participants appeared to agree should be adopted. This standard can be difficult for data entry as people prefer using two separate columns for date and time. A solution is to incorporate a translator and to adapt the standard to allow two columns.

Coordinates: Many coordinate systems were mentioned but none were agreed upon as a best-practice. The original coordinates, coordinate system and accuracy were suggested as important metadata.

Missing values: The different types of missing values need consistent codes or at least be defined in the metadata. Standards could be based on what has been already developed by the NTGS.

Ground temperature file structure: Participants agreed that both the 'wide' and 'long' format of ground temperature data are useful at different stages of data processing. A wide-to-long translator tool was suggested as a possible contribution to aid data management.

Participants also discussed developing a set of minimum metadata requirements. It was noted that these may be difficult to mandate and would be impossible to follow for legacy data. The metadata that were discussed in the session included:

- Accuracy and precision of measurements
- Classification of quality / level of processing
- Dominant vegetation
- Backfill materials
- Field photos
- Ground surface elevation
- Missing value code definitions

Participants appeared to agree that guidelines are needed to have standardized data. Although no specific consensus was reached for how this should be done, several ideas were generated:

1. Provide instruction manuals

2. Provide ‘gold standard’ examples of data structure and included metadata for different data types (as done by Integrated Ocean Observing System)
3. Provide a central resource for guidelines for best practices and conventions, both in terms of data sharing and in terms of field techniques
4. Create a validation tool that checks compliance with a set of guidelines (as done by the Arctic Data Center)

7) Taking Action

Summer of data

Due to the current COVID-19 restrictions, fieldwork this summer will be significantly reduced or cancelled entirely. While unfortunate, this does create an opportunity to make significant progress on data-related activities. The purpose of this session was to scope these activities and lay out tangible next steps. Three themes were identified based on previous workshop discussion: gathering and processing legacy data, working towards a glossary in collaboration with the CPA, and liaising with Northern partners to coordinate possible data collection.

Participants chose a theme and were asked to identify what needs to be done, what support is needed or available, and as a minimum output, three actionable items that can take place within the next several months.

For the harvest of legacy data, the four stages of the work that were identified were: determining where to find the data, creating an inventory of what has been done, understanding what is needed by the data users and describing the data quality or qualities. To support this, one or more individuals are needed with an understanding of what data are out there and, in addition, personnel to do some of the hard work of gathering and cleaning it, possibly including network HQP. As next steps, participants identified: creating a sub-committee of people to take up these tasks, coming up with a strategy, and establishing a better understanding of user needs – possibly though a survey.

Updating the permafrost terms glossary has been proposed as an action group activity by the CPA. This includes the development of a glossary that can grow and develop, that integrates other permafrost-related disciplines, and is not limited to a static pdf document. Based on discussions on day 2 of the workshop, connections in PermafrostNet were seen as a way to connect the glossary to the semantics community. To support this work, someone would need to lead the development of the glossary and integrate with semantic working groups. A postdoc was suggested as one possibility. In addition, a panel of experts to provide the scientific consensus would also be required. The next steps for this are to: identify a person or group to lead the effort and to develop a document outlining the framework of the glossary. The group of participants suggested that PermafrostNet take an active role in working with the CPA to kick off the project.

Because of restricted travel to the territories this summer, Northern network partners may be able to assist with certain field work elements that cannot be completed by those currently living in the South. For this to happen, relationships with Northern communities and organizations must be built and strengthened. The importance of long-term relationships was brought up as a way to mitigate risks to long-term environmental monitoring from large-scale events such as COVID-19. Resources must also be made available to support this work. This may include compensation, equipment, and the development

of physically-distanced field protocols. In the short-term, participants suggested that the workshop themes be relayed to community members to let them know their input and help is valued. Fostering new partnerships and strengthening existing ones was another suggestion. Environmental monitors were identified as a possible group who could assist with this work, with the understanding that funding would be required and they would expect to return to their usual duties after the outbreak is over.

Long-term strategy

To help guide the long-term development of a permafrost data system, participants were asked to imagine themselves in 10 years using such a system, from the perspective of a someone accessing the data, someone submitting data, and as someone involved in the governance and decision making. The majority of participants cycled between breakout rooms, while one host stayed behind to take notes and synthesize ideas between groups.

From the perspective of a data user (Table 8), the imagined system is one that makes data easy to find, and accessible to experts and non-experts alike. This is achieved through providing several ways to access the data, which may include an API, an interactive interface, and a way to get detailed information about a site or dataset. In addition to the raw data, value-added products are also available. The data system is interoperable with other research domains or sources, and has rich metadata to detect duplicate datasets, or provide more information in publications.

Table 8: Elements of a permafrost data system 10 years from now, from the perspective of a data user.

Accessing data	
Interoperable	• Links to cross disciplinary impacts and other database not related to permafrost • Standards and methods to mesh datasets from different sources • Data from multiple sources accessible whichever database is used.
Findable and Accessible	• Data are as easy to find as academic papers • Datasets easily searchable and harvestable from generic search engines • Catalogue service that offers searchable, findable metadata
Useable by non-experts	• Online interface that doesn't require specific skills • Interactive online for less technical users to synthesize knowledge
Multiple ways to access	• Drill down in individual data sets • Both downloading data and bringing models to data online • Flexible use-cases • Have an API for programmatic access
Provides data products	• Analogue reports to store the turn data into information & knowledge • Interpretations of the data available • Visuals and reports for the area built in the metadata
Referenced	• Links to the research papers that were published • Searchable – DOI • Duplication alerts

From the perspective of someone adding their data to the system, (Table 9), the process of submitting data is simple and clear. They have the support of someone at the data centre to help with certain elements, and there is a division of responsibilities between the data creator and the data manager. Tools for data submission and cleaning are available. Quality control elements are split between manual and automated components. The system itself hosts historic datasets that can integrate with new data.

The system is decentralized; the data creator may submit their data to someone locally responsible, but it will be visible in the larger system.

There were some contradictory suggestions in this session, namely about whether data should be quality checked automatically, or manually by a person, and what the balance of responsibility should be between the field scientist and data curator. On the one hand, one suggestion was that once the data are collected, all further processing should be done by the data manager. The other perspective is that the data creator is the one most familiar with the data and should therefore be the one to ensure it is properly checked and correct.

Table 9: Elements of a permafrost data system 10 years from now, from the perspective of a data creator.

Submitting data	
Tools for data submission and curation	• Data and metadata checkers highlight missing info, or deviation from standards • Visualisation, transformation and processing included in the submission process • Users can upload in their own format and transformations are performed automatically • Automated QA/QC tools
Don't reinvent what exists	• Collaboration is necessary - use existing database • Don't duplicate efforts and avoid overlap • Avoid rework as much as possible so as to not recreate what's been done before
People to help and a balance of responsibilities	• Metadata input from the person that collected it • Key contact person is available to assist • A person performs QA/QC on the data • Responsibilities of data collector and data manager are clear
Simple process	• Make submission as automated and simple as possible, non-time consuming • Simple field collection/templates (on tablets?) • The process is very easy
Decentralized	• Data generators and data collection should be de-centralized • Community partners are provided with the resources to collect and enter data
Standards	• Data itself needs to be comparable to previous data (backwards-compatible) • Metadata templates with defined vocabulary • Template for data collection

From the perspective of someone involved in the management or governance of the data system (Table 10), they work with a diverse group representing stakeholders from multiple sectors and user groups. The data system has a clear mandate which includes both the provision of data and the creation of summary data products. There is strong government involvement to ensure stable funding; due to the decentralized nature, this involves governments at multiple levels. The governance structure is flexible enough to respond to changes in the needs put upon it.

Table 10: Elements of a permafrost data system 10 years from now, from the perspective of someone involved in the governance body.

Managing data	
Strong government involvement	• Needs long-term support and stable funding • Academia / NSERC are not in the business of curating data

Diverse representation	• Representatives from user groups • Take advantage of differences between sectors: academia can provide innovation and nimbleness while government provides stability • Indigenous, territorial & federal governments, academia, industry
Clear Mandate	• Identify gaps, find a niche, and be clear about what kinds of data are within that mandate • Requirements for engagement between organizations for use and maintenance • Data curation as part of requirements • Active use of data through creation of annual summary reports, data products
Decentralized	• Regional structure that feeds into national and international committees • Under a federal umbrella with regional management • Funding at national level for coordination between regions • Strong coordination between nodes and with international community
Flexible	• Governance structures should be flexible to external and internal changes

The final part of the discussion focused on how to communicate the benefits of funding a data management system as a foundational piece for permafrost research and engineering. Discussion was focused on working towards **data management as an essential part of any project or contract:** currently, most permafrost data management is done using soft money through academic project funding or employees on fixed-term contracts. Because of its importance for monitoring and engineering design, the focus should be on securing more stable funding, for example through existing infrastructure development projects. In the case of large northern infrastructure projects, a relatively small fraction of this investment would significantly benefit northern capacity for permafrost data management.

Other data-providing organizations were discussed as possible models such as the Water Survey of Canada and the Canadian Integrated Ocean Observing System. Participants noted that these likely grew out of the demand for the data. Despite permafrost issues being ubiquitous in the North, governments may not yet appreciate the relevance of permafrost data.

Finally, the **impact of government COVID-19 spending** on the sustainability of permafrost data management was seen to present **opportunities as well as challenges**: There were mixed opinions on how the recent government spending would affect the long-term sustainability of any permafrost data management system that develops in the next 5 to 10 years. On the one hand, concerns were raised that research was classified a *non-essential service* and that funding cuts may disproportionately affect the resources that might otherwise be made available for the development of a permafrost data management system. On the other hand, continued spending on Northern infrastructure was seen as one of the ways the government may mitigate a possible recession because it remains a key driver of the economy. The case for better funding for permafrost data management is only strengthened by increased infrastructure spending and climate change.

There was some divergence of opinions on how this uncertainty should be handled. One proposed solution was to rely as much as possible on existing international data systems to ensure long-term stability and minimize risk. Other participants felt that interfacing with and benefiting from other international efforts was important, but that Canada has significant needs for which it should not rely on others providing the services.

Motivating progress:

On the last day, a group of participants set out to answer the question: "*How do we motivate individuals or teams to take on the often large and unrewarding task of uploading current and historical data?*". The group self-organized to come up with a framework, laying out steps to take in the short- and long-term. Motivation was seen to occur at two levels: at the strategic level where direction is established and at the operational level where work gets done.

The ideas discussed for motivating proper data management and data sharing can be summarised into five themes: (1) **Resources**, which are needed to make things possible; (2) **Recognition** of those involved in each step of the data creation tasks. This can be achieved through references, and at the network level through an annual database report which highlights top data contributors. Keeping track of data usage and having data users provide feedback can also be motivating; (3) **Communication** with and among those completing the data management tasks is very important to motivate at the operational level. Annual workshops would facilitate sharing methods and challenges and highlighting the outcomes and final products. A working group of people involved in data management could promote continuous communication between annual events; (4) A **governance structure,** with a leader to mobilize people, and a committee with a cross-section of stakeholders. With leadership and clear objectives, operational measures are implemented that align with the long-term data management strategies. (5) The adoption of **policies or regulations**, with requirements to incorporate data management into projects or to upload data in a standardized format and within a certain time frame of collection. Territorial Governments have already started writing these requirements into their contracts.

Table 11: Participant comments on how to motivate progress for data-related activities

	Next steps	**Short-term**	**Long-term**
Secure Resources	• Identify what resources are required (and available!)	• Develop common automated methods with documentation	• Stable funding for sustainability
Create governance & leadership	• Establish leaders and high-level structure	• Create working groups with regular meetings to share experience • Increase capacity for legacy data	• Need a strong strategic leader or champion
Give recognition		• Recognize people doing the work • Track data usage, so we can see that our efforts are being put to good use • Demonstrate the final products so people can see what they are contributing to	

Communicate		• Involve the private sector (this may require additional resources) • Unify the fragmented data landscape, including within the public sector	• Annual workshop to talk about what data you are looking for to incentivise collaboration and efficiency in data collection. • Develop connections between data collectors, data entry and end users
Change regulation & policies		• Create accountability through data management plans in new project proposals	• Regulatory or contract requirements for data submission

8) Final thoughts and next steps

The outputs of this workshop point to a need for continued collaboration and communication at all steps of the data creation process. NSERC PermafrostNet is currently organizing semi-regular meetings for those involved in developing permafrost data systems in Canada. These will continue to help ensure data systems do not develop in isolation, and to strengthen the Canadian permafrost data community of practice. Increasing involvement from the private sector is a priority to diversify representation and also help advocate for the need for better permafrost data management as a part of new projects or contracts. The next planned formal opportunities to discuss permafrost data will be at an online session as part of the NSERC PermafrostNet AGM in November 2020, and at the *permafrost data systems* workshop session at the *Regional Conference on Permafrost* (RCOP) in Boulder, Colorado in July 2021[10].

Some of the issues brought up during the workshop highlighted how field data collection methods can impact the ease with which data can be shared and disseminated. Although these issues were out of scope for this workshop, NSERC PermafrostNet had planned a fieldwork rehearsal to harmonize and share data collection methodologies which could have been a forum to discuss some of them. This event was scheduled to be held in the days leading up to the data workshop, and the two events were intended to complement one another. A similar event in the future, possibly combined with a smaller discussion on data management, may offer an opportunity to revisit this area of overlap.

Going forward, the next steps for NSERC PermafrostNet will include releasing version 1 of the network data policy, and developing a white paper outlining a vision for permafrost data in Canada, based in part on the outcomes of this workshop. The network will also continue to prototype integrated permafrost data management and shared governance. The network provides a forum that is fundamental for data stewardship and decision making so that researchers can focus on research. By connecting permafrost data management expertise (GSC, YGS, NTGS and elsewhere), a system can be developed that is interoperable and sustainable. For more information, or to get involved, contact Nick Brown (nick.brown@carleton.ca) the Permafrost Data Scientist of NSERC PermafrostNet or visit the PermafrostNet website (permafrostnet.ca/data).

[10] https://www.uspermafrost.org/21rcop/sessions.shtml

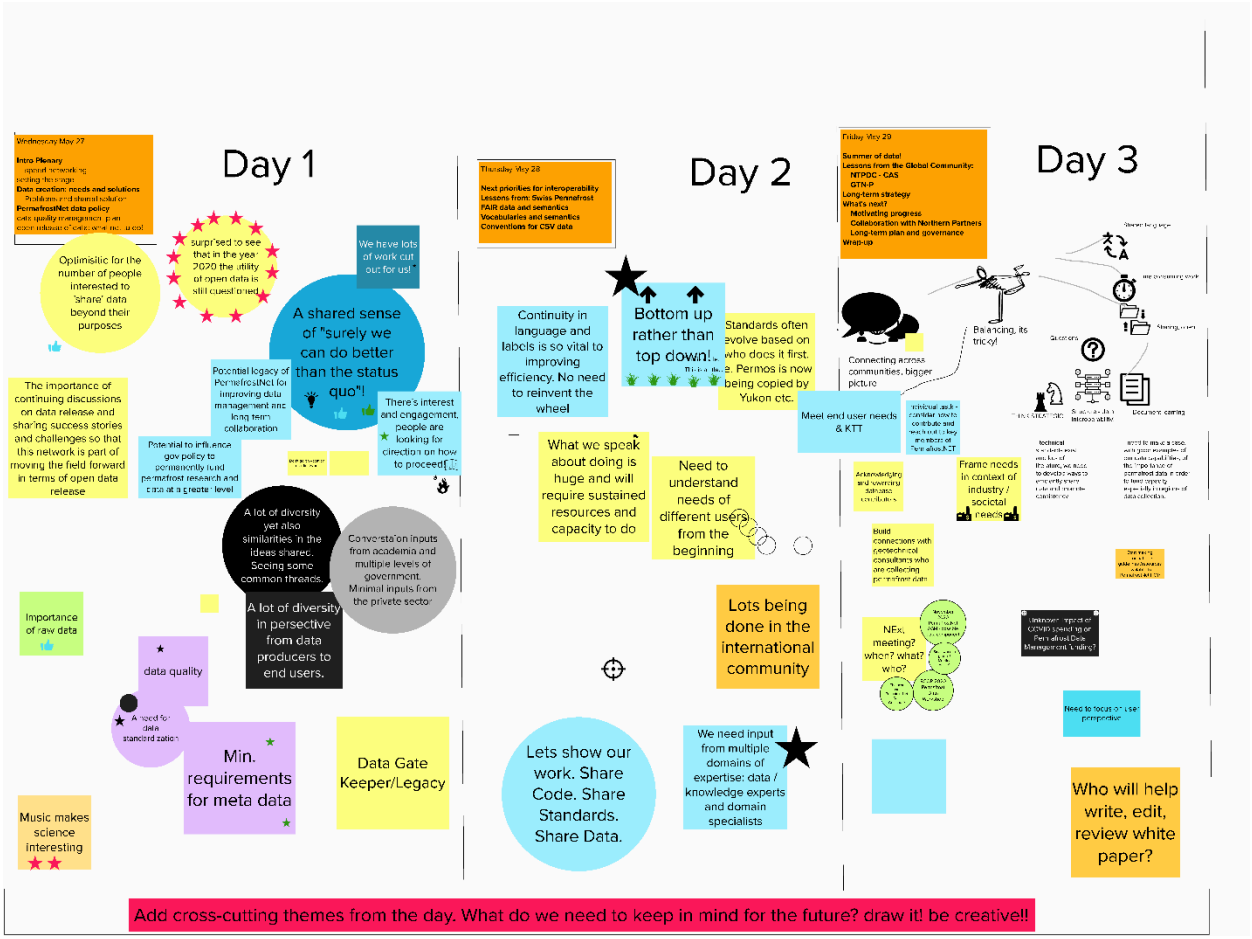


Figure 9: At the end of each day, participants were asked to jot down ideas or cross-cutting themes on a Mural board that synthesized the results from the day. A full-resolution version of this board is provided inAppendix IV.

9) Appendices

Appendix I: Workshop agenda

	Wednesday 27 May
10:00 - 10:30	*(Optional) coffee & networking*
10:30 - 11:45	Welcome & Plenary Session: Challenges & Opportunities for the Permafrost Community
11:45 - 12:00	*Coffee / snack break*
12:00 - 13:15	Data creation and access: What's needed for ground temperature and basic geotechnical data
13:15 - 13:40	*Lunch break*
13:40 - 14:00	*Lunch-and-learn: Discover what permafrost data are available in Canada from the GSC, NWT Geological Survey, and YGS*
14:00 - 15:15	PermafrostNet data policy: open release and quality control
15:15 - 15:20	*Break*
15:20 - 15:35	Closing - tie-ins for white paper

	Thursday 28 May
10:00 - 10:30	*(Optional) coffee & networking*
10:30 - 10:35	Recap from day 1
10:35 - 11:10	Beyond ground temperature: prioritizing permafrost data types for future interoperability
11:10 - 11:45	Plenary:
	a) Lessons from the international community part I
	b) background on data interoperability and semantics.
	Guest speakers Jeanette Noetzli & Peter Pulsifer
11:45 - 11:55	*Coffee / snack break*
11:55 - 13:10	Speaking the same language: vocabularies and semantics for permafrost
13:10 - 13:55	*Lunch break*
13:55 - 15:10	Adopting standards and specifications for network data
15:10 - 15:15	*Break*
15:15 - 15:30	Closing - tie-ins for white paper

	Friday 29 May
10:00 - 10:30	*(Optional) coffee & networking*
10:30 - 10:35	Recap from day 2
10:35 - 11:10	Summer of data: how to make the best of the field season during a pandemic

11:10 - 11:40	Plenary: Lessons from the international community part II.
	Guest speakers Xiaoduo Pan & Dmitry Streletskiy
11:40 - 11:55	*Coffee / snack break*
11:55 - 13:10	Long-term strategy: what is needed for permafrost data in Canada?
13:10 - 13:55	*Lunch break*
13:55 - 14:55	Your priorities: what needs more discussion? (3 time slots available, topics to be determined)
14:55 - 15:00	*Break*
15:00 - 15:30	Closing - next steps

Appendix II: Participant list

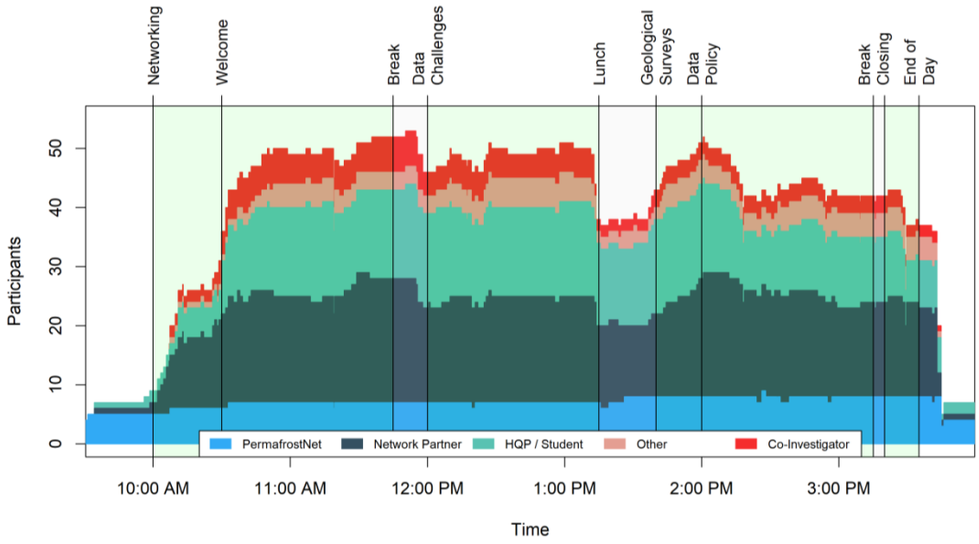
Name	**Location**	**Organization**
Tristan MacLean	Ottawa, ON	PermafrostNet
Nick Brown	Ottawa, ON	PermafrostNet
Shirley McKey	Ottawa, ON	PermafrostNet
Anne Barker	Ottawa, ON	National Research Council
Alex Gao	Ottawa, ON	PermafrostNet
Michel Paquette	Kingston, ON	Queen's University
Emma Stockton	Ottawa, ON	PermafrostNet
Danielle Chiasson	Churchill, MB	Churchill Northern Studies Centre
Hannah Macdonell	Ottawa, ON	Carleton University
Catherine Kim	Ottawa, ON	Transport Canada
Alex Bevington	Prince George, BC	BC Government
Moya Painter	Yellowknife, NT	NWT Geological Survey
Stephan Gruber	Ottawa, ON	Carleton University
Jenna Craig	Ottawa, ON	Transport Canada
Michelle Blade	Iqaluit, NU	Canada-Nunavut Geoscience Office
Igor Egorov	Ottawa, ON	National Research Council
Dawn Rybchynski	Yellowknife, NT	Government of the NWT
Ashley Rudy	Yellowknife, NT	NWT Geological Survey
Dima Streletskiy	Washington	George Washington University
Pascale Roy-Leveillee	Sudbury, ON	Laurentian University
Michelle Laurie	Rossland, BC	PermafrostNet
Sharon Smith	Ottawa, ON	Geological Survey of Canada
Jocelyn Hayley	Calgary, AB	University of Calgary
Samuel Gagnon	Montreal, QC	Université de Montréal
David Arthurs	Ottawa, ON	PolarView
Tim Ensom	Yellowknife, NT	Wilfred Laurier University
Zakieh Mohammadi	Calgary, AB	University of Calgary

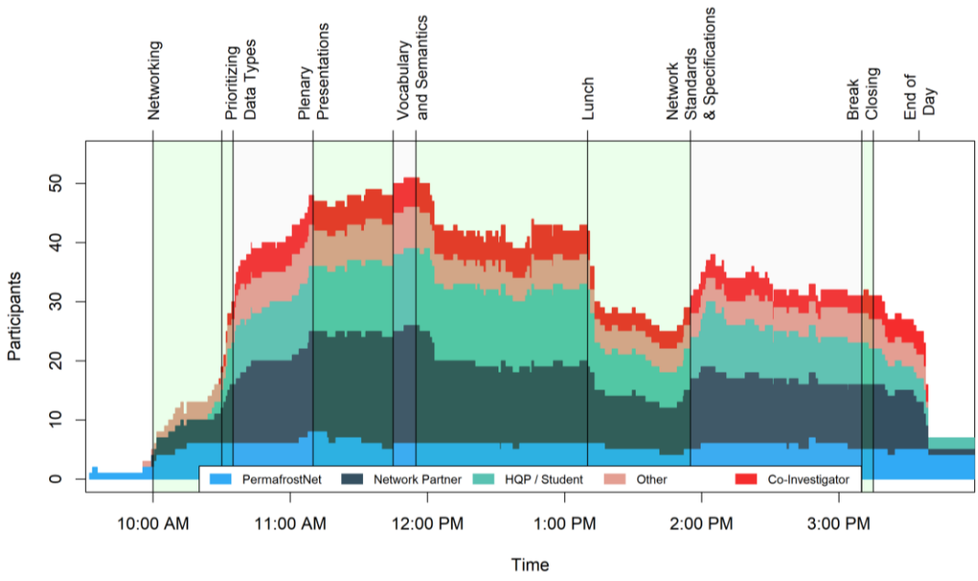
MohammadHossein Gamshadzaei	Montreal, QC	Université de Montréal
Linda Ham	Iqaluit, NU	Canada-Nunavut Geoscience Office
Brendan O'Neill	Ottawa, ON	Geological Survey of Canada
Alice Wilson	Inuvik, NT	Government of the NWT
Sara Brown	Yellowknife, NT	NWT Association of Communities
Daniel Fortier	Montreal, QC	Université de Montréal
Joe Melton	Victoria, BC	Environment and Climate Change Canada
Panya Lipovsky	Whitehorse, YT	Yukon Geological Survey
Adam Kirkwood	Sudbury, ON	Laurentian University
Jonas Darey	Montreal, QC	Université de Montréal
Sarah Gauthier	Quebec, QC	Université Laval
Bingqian Zhang	Vancouver, BC	Simon Fraser University
Emilie SJ	Whitehorse, YT	Carleton University
Christina Béland	Iqaluit, NU	Qaujigiartiit health research centre
Patrick Jardine	Ottawa, ON	Carleton University
Shawn Kenny	Ottawa, ON	Carleton University
Brian Sieben	Ottawa, ON	Environment and Climate Change Canada
Kate Swan	Ottawa, ON	Carleton University
Eva Thévenin	Montreal, QC	Université de Montréal
Miki Ehrlich	Yellowknife, NT	NWT Association of Communities
Matt Jones	Santa Barbra	Arctic Data Center
Christopher Burn	Ottawa, ON	Carleton University
Bin Cao	Beijing	Chinese Academy of Sciences
Kala Pendakur	Ottawa, ON	Standards Council of Canada
Cameron Ross	Kingston, ON	Royal Military College
Jeannette Nötzli	Davos	PERMOS
Fereshteh Ghiami Shomami	Gifu, Japan	Gifu University
Federico		
Elisabeth Hardy-Lachance	Montreal, QC	Université de Montréal
Xiaoduo Pan	Beijing, China	Chinese Academy of Sciences
Brett Elliot	Whitehorse, YT	Yukon Geological Survey
Antoni Lewkowicz	Ottawa, ON	University of Ottawa
Kate Grandmont	Montreal, QC	Université de Montréal
Peter Pulsifer	Ottawa, ON	Carleton University
Jarrod Haas	Vancouver, BC	Simon Fraser University
Jan Beutel	Zurich, Switzerland	ETH Zurich
Maribeth Murray	Calgary, AB	CCADI, University of Calgary
Alessandra Kempson	Sudbury, ON	Laurentian University
Kati Laakso	Edmonton, AB	University of Alberta

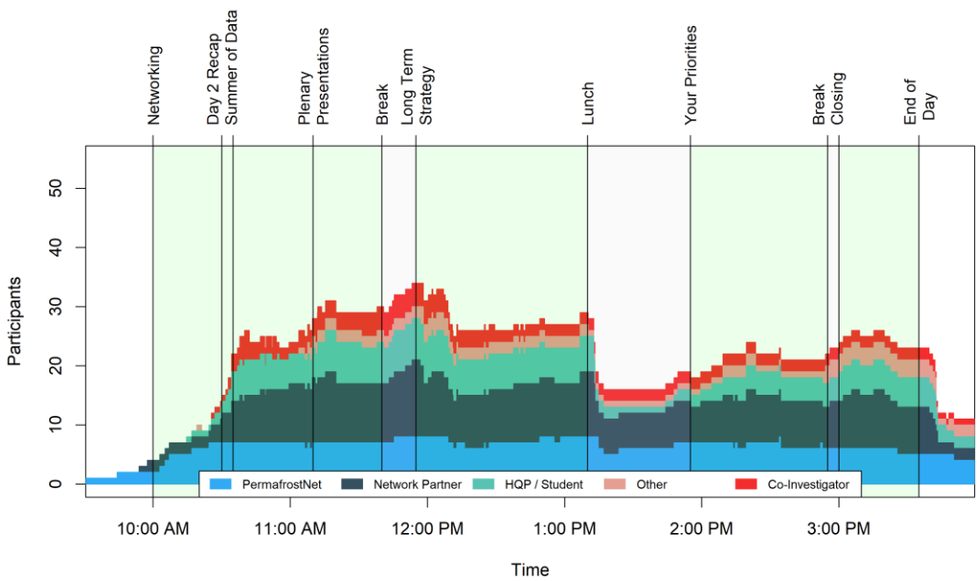
Ravi Darwin Sankar	Calgary, AB	CCADI
Danika Ouellette	Calgary, AB	University of Calgary
Boyan Brodaric	Ottawa, ON	Natural Resources Canada
Beth Cowan	Yellowknife, NT	Government of the NWT
Logan Rudkevitch	Yellowknife, NT	Government of the NWT
Manny Kudlak	Sachs Harbour, NT	Inuvialuit Game Council
Etienne Godin	Quebec, QC	Université Laval
Srini Sundaram		
LeeAnn Fishback	Churchill, MB	Churchill Northern Studies Centre
CORIS	Ottawa, ON	Carleton University
Steve Kokelj	Yellowknife, NT	NWT Geological Survey
Lin Chen	Quebec, QC	Université Laval

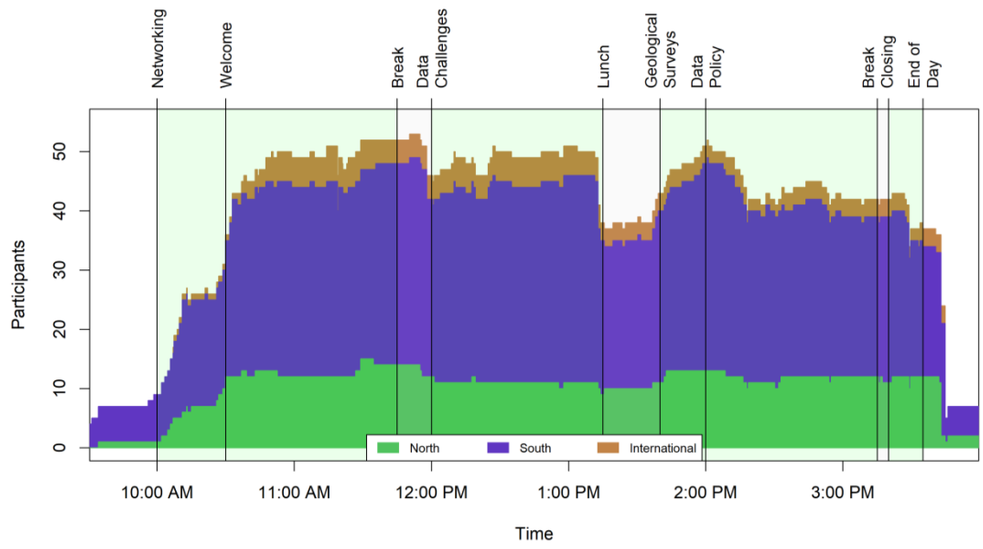
Appendix III: Participation throughout the day

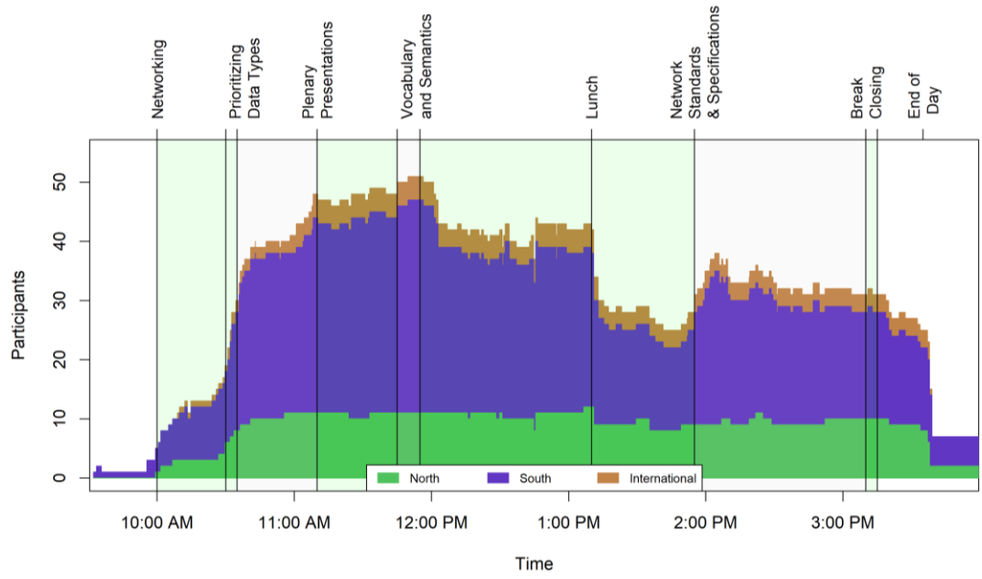
The level of attendance throughout the day was obtained from the zoom records. After cleaning the data to account for participants renaming themselves, the total level of attendance was plotted. Participants were classified based on their affiliation to the network and based on their location (Northern Canada, Southern Canada and International)

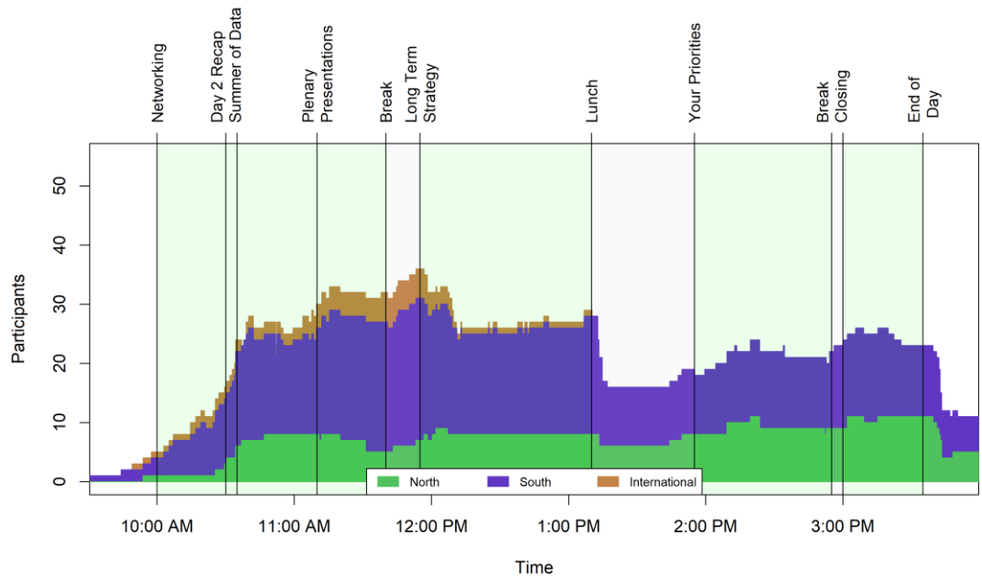







Networking
Welcome
Break
Data Challenges
Lunch
Geological Surveys
Data Policy
Break
Closing
End of Day
Participants
0
10
20
30
40
50
North
South
International
10:00 AM
11:00 AM
12:00 PM
1:00 PM
2:00 PM
3:00 PM
Time


Networking
Prioritizing Data Types
Plenary Presentations
Vocabulary and Semantics
Lunch
Network Standards & Specifications
Break
Closing
End of Day
Participants
0
10
20
30
40
50
North
South
International
10:00 AM
11:00 AM
12:00 PM
1:00 PM
2:00 PM
3:00 PM
Time


Networking
Day 2 Recap Summer of Data
Plenary Presentations
Break
Long Term Strategy
Lunch
Your Priorities
Break
Closing
End of Day
Participants
0
10
20
30
40
50
North
South
International
10:00 AM
11:00 AM
12:00 PM
1:00 PM
2:00 PM
3:00 PM
Time

Appendix IV: Mural boards

From your perspective, what are the challenges you face in data creation, handling and sharing?

Data Fusion across different sources (e.g. satellite imagery, borehole, UAV, strain gauge)

instructions to help new people get into routine data tasks, import, cleaning,....

TIME!

Track service needs to field sites as part of data base

Publish raw and cleaned data with field notes, get DOI, then insert into database – currently no good solution

new naming, different acronyms, new structure every time (different vocabulary for same concepts)

large technical learning curve (databases, scripting etc)

Automated scripts for repetitive tasks i.e. appending and QA/QC of annual ground temperature downloads - BONUS Scripts for all different sensors!

reinventing the wheel every time you collect/create data

erratic project partners not wanting to share data causing huge aggravation

Lots of obstacles to making data available to the puclic (confidentiality issues, firewalls...)

get support from higher level government / partners

I am sure there are a number of individual solutions by individual researchers but would be nice to share and possibly simplify work flow

Being a student and formatting data so that it's useable for more than just my personal project, and can be formatted for further use by others after project completion

I'm not sure how to best format my data

I need help cleaning up my data

Static methods and manual collect on the field makes it hard to have regular updates and makes it hard to have an efficient data flow

I need an efficient way to remove sensor dropouts from my measurements

I never miss fieldwork, but then back in the office hard to make time for data management.... gets pushed to the bottom of the list

Data collection

making good data handling practice part of my teaching and supervision needs insane amounts of time. Need help.

no metadata, or little/inconsistent metadata

Data cleaning and consolidation is very time consuming and repetitive

sensor drift, damaged by animals, outliers

software for (semi) automated warning flags

read up/study data science basics

collaborate with colleagues / find a mentor

There is a project at the CEN about interactive permafrost data visualisation in Nunavik with ArcGIS online Operation dashboard

interactive visualization of data

r-shiny-server for data viz

ArcGIS Online

Problem: I don't know where to put my data to make it available

Harmonizing different field observation variables and methods

sharing solutions and workflows publicly (github etc)

lack of mathematical/physical understanding with data collectors

"Gold standard" examples of how to structure datasets with explanation (requires developing /adopting some of these standards)

limited staff resources/lack of continuity

how to assess/classify legacy data quality

Use consistent vocabularies for variables and methods

Arctic Data Center Semantic Measurement Search (M Jones)

Using similar data formats so that tools can be reused between research groups

data redundancy through 'geo-stations' collecting multiple types of data, and knowing other conditions

Managing large volumes of satellite data for permafrost feature extraction

using image services / web mapping

The Permafrost Network has access to the CCADI/Polar TEP platform with access to petabytes of EO data and tools for its analysis.

I use Nordicana-D

Determining when to release data vis a vis competing priorities like pre- or post-publication, impact on student projects etc.

As a student I have problem having access to available data. Making previous data more researchable and accessible by compiling them in one comprehensive database can be a solution.

Regarding the data I am going collect during my project having standard format make my date useful for future researchs.

Adopt shared vocabularies - community buy-in and scientific consensus

Funding/project starts and stops.

Limited resources to support data mgmt

We need sentinel sites that will be maintained in the long run

The NTGS and GSC have a number of Sentinel sites throughout the NWT that are maintained and downloaded annually. The NTGS is working to expand this network outside of infrastructure corridors (historically a lot of focus in these areas)

How many more are needed

*What existing tools and services would solve the problem? (**GREEN**) Add your name so we can get in touch later too!*

*What possible or futuristic tools and services may solve the problem? (**BLUE**) Add your name so we can get in touch later too!*

PermafrostNet creating guidelines to students working in the network on formatting data consistently throughout the network (perhaps also creating data repository? -Adam K

build a separate data repository for the network

Work with Arctic Data Committee on Federated data search (P Pulsifer)

Existing repository federations: DataONE (M Jones)

Provide a location to host our raw data -quickly at the end of the fieldseason, then we can take our time to clean/update

Existing open data repositories: pangea, github, http://www.opendatarepository.org/, zenodo

Existing repositories: Arctic Data Center (M Jones)

Most data that is being discussed comes from individual projects. The GSC, in particular, has locations that are maintained. Universities, with a very few exceptions will not do this in teh long run

Surely the existing data bases, GSC, NTGS etc have an established format. Why reinvent the wheel?

Or can we ensure that we share data in a way that is accessible for the general user, i.e. communities, public, but also has the depth and metadata required for use by an expert

USER STORIES! Describe the requirement based on the context of who needs it any why.

I'd like to search for similar field locations (near mine), with some contact information so that we can reach out and collaborate.

i.e. metadata catalogue/map of canadian permafrost data

Not sure where to put this!

We could create a standards document or guidelines for how to properly structure / clean data

Provide higher level summary data for less technical users/decision makeers/consultants

It would be useful for new students coming in to have resources available to help format or clean their data, and also general guidelines on how to provide informative metadata.

To demonstrate the value of the network to communities, partners, and Canada, we need to ensure that the data is accessible, shared in a timely fashion and user friendly

also need to share 'success' stories

As a database and interface developer, I need to understand the use cases of the users of the data. One suggestion was the idea of ensuring that every data download comes with a fully described field-by-field comment file with acronyms expanded, units... Dawn Rybchynski if you want to tell me your use case.

The comment that some people need the outliers really helped as well. It may affect how the cleaning phase happens and what we consider to have as "Raw". i.e. a broken sensor may need to be cleaned but maybe it should stay because someone may be doing a reliability study and want to see that failure data.

Another good point from one of the presentations is that without a standardization of observational/field methods and tools data standardization is always going to be tricky. We won't be able to get away from some measure of scripting to 'make it all the same'.

Day 1

Wednesday May 27

Intro Plenary
speed networking
setting the stage
Data creation: needs and solutions
Problems and shared solution
PermafrostNet data policy -
data quality management plan
open release of data: what not to do!

Optimisitic for the number of people interested to 'share' data beyond their purposes

surprised to see that in the year 2020 the utility of open data is still questioned

We have lots of work cut out for us!

A shared sense of "surely we can do better than the status quo"!

Potential legacy of PermafrostNet for improving data management and long term collaboration

There's interest and engagement, people are looking for direction on how to proceed!

The importance of continuing discussions on data release and sharing success stories and challenges so that this network is part of moving the field forward in terms of open data release

Potential to influence gov policy to permanently fund permafrost research and data at a greater level

Don't go on vacation after fieldwork.

A lot of diversity yet also similarities in the ideas shared. Seeing some common threads.

Converstaion inputs from academia and multiple levels of government. Minimal inputs from the private sector

A lot of diversity in persective from data producers to end users.

Importance of raw data

data quality

A need for data standardization

Min. requirements for meta data

Data Gate Keeper/Legacy

Music makes science interesting

Day 2

Thursday May 28

Next priorities for interoperability
Lessons from: Swiss Permafrost
FAIR data and semantics
Vocabularies and semantics
Conventions for CSV data

Continuity in language and labels is so vital to improving efficiency. No need to reinvent the wheel

Bottom up rather than top down!

This is a title...

This is a title...

Standards often evolve based on who does it first. e. Permos is now being copied by Yukon etc.

What we speak about doing is huge and will require sustained resources and capacity to do

Need to understand needs of different users from the beginning

Lots being done in the international community

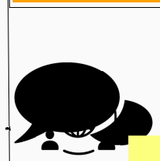
Lets show our work. Share Code. Share Standards. Share Data.

We need input from multiple domains of expertise: data / knowledge experts and domain specialists

Day 3

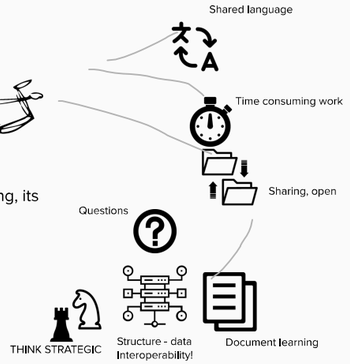
Friday May 29

Summer of data!
Lessons from the Global Community:
NTPDC - CAS
GTN-P
Long-term strategy
What's next?
Motivating progress
Collaboration with Northern Partners
Long-term plan and governance
Wrap-up



Connecting across communities, bigger picture

Balancing, its tricky!



Meet end user needs & KTT

Individual tastk - consider how to contrbute and reach out to key members of PermafrostNET

Acknowledging and rewarding database contributors

Frame needs in context of industry / societal needs

- technical standards exist and lots of literature, we need to develop ways to efficiently share data and promote consistency

- need to make a case, with good examples of our data capabilities, of the importance of permafrost data in order to fund capacity especially in regions of data collection.

Build connections with geotechnical consultants who are collecting permafrost data

Start making information / guidelines /resources available to PermafrostNet HQP

NExt meeting? when? what? who?

November 2020 PermafrostNet AGM - possible data component

Small working groups? Meeting online?

The next annual PermafrostNet Data Workshop?

RCOP 2020 Permafrost Data Workshop

Unknown Impact of COVID spending on Permafrost Data Management funding?

Need to focus on user perspective

Who will help write, edit, review white paper?

Add cross-cutting themes from the day. What do we need to keep in mind for the future? draw it! be creative!!