WORLD METEOROLOGICAL ORGANIZATION

GLOBAL CRYOSPHERE WATCH

REPORT No. 19
(2017)

GCW Data Portal Team meeting
Fourth session

Montreal, Canada, 14-15 September, 2017

GCW Technical Report # 19 (2017)

World Meteorological Organization
Weather · Climate · Water

GCW
Global Cryosphere Watch

Meeting venue:
Room André-Robert, Environment and Climate Change Canada, Place Bonaventure (800 de la Gauchetière Ouest, suite 7810, on the 7th floor), Montreal.

Chair, Publications Board
World Meteorological Organization (WMO)
7 bis, avenue de la Paix
P.O. Box 2300
CH-1211 Geneva 2, Switzerland

Tel.: +41 (0) 22 730 8403
Fax: +41 (0) 22 730 8040
E-mail: Publications@wmo.int

NOTE

Executive Summary

The GCW Data Portal meeting took place on 14 and 15 September 2017, in Montreal being hosted by Environment and Climate Change Canada. The meeting was opened by Dr Barry Goodison, Vice-Chair of the Global Cryosphere Watch (GCW) Steering Group (GSG) and Dr Øystein Godøy, the Chair of the GCW Data Portal Team. They noted that the goals of the meeting : (1) GCW Operational Manual for contributing data centres and GCW Guidance for data centres contributing to GCW; (2) road map for developing and testing the GCW Interoperability Package for CryoNet stations, (3) requirements for compliance to WIS and WIGOS, and (4) road map for developing standard vocabularies for metadata for the cryosphere variables of the observing programs of GCW (CryoNet). The meeting was attended by members of the team, as well as representatives of several CryoNet stations and partners.

It was noted that the GCW Data Portal is a mechanism to provide access to relevant cryosphere datasets, and is developed as a Distributed Data Management system, metadata driven, net centric, and it will be a WMO Information System (WIS) Data Collection and Production Centre (DCPC), and compliant to the WMO Integrated Global Observing System (WIGOS) regulatory framework. The participants acknowledged that the OAI-PMH protocol is the interface for exchanging metadata with the GCW Portal, and that the metadata has to be provided in ISO19115 and/or GCMD DIF formats.

It was noted that many CryoNet stations don't have data management systems. The success of interoperability of the Data Portal and each CryoNet or contributing station requires support from GCW for e understanding and applying the WIGOS, WIS requirements. A working Change Management System will be implemented to track and communicate endpoint changes.

The participants reviewed the GCW Interoperability Package for CryoNet stations for coupling the GCW Data Portal with a small data centre. It was agreed that Kluane Lake CryoNet Station, as well as Environment and Climate Change Canada DCPC and Sodankylä CryoNet site will be engaged, and the CryoNet stations operated from the Chinese Academy of Science will be kept in the loop.

The participants noted the need to develop controlled vocabularies with definitions broadly accepted by the scientific communist, for the discovery of metadata and data, to address the current gaps. The participants expressed support for the initiative of GCW to develop a set of key words and their associated definitions, and to promote with Data Centres the use of the recommended keywords. It was noted that this is a goal of many other communities, and collaboration is strongly encouraged.

Peter Pulsifer, the Chair of the Arctic Data Committee of SAON introduced the concept of the polar data forum, planned for 2018, for developing a statement for the Arctic Science Ministerial (2018) promoting the data standardization and interoperability at system level, and engaging the scientific and operational communities active in the arctic. GCW accepted to be an active contributor in the organization of the forum, building on the work of the Data Portal.

Feiteng Wang provided an overview of the cryosphere observing stations of the Key State Laboratory of Cryosphere Sciences of the Chinese Academy of Science, which include five CryoNet stations. It was agreed that the interoperability between the data management system used for these stations and the GCW Data Portal is a high priority work.

The participants agreed to follow up on the topics resulting from this meeting at the Arctic Change Conference, in Quebec City, in December 2017, and that a side event on cryospheric data exchange will be organized at EC-70, June 2018.

All meeting actions are included in Annex 4 of this report.

The meeting ended at 16:00 on 15 September 2017.

Contents

1. ORGANIZATION OF THE SESSION (Ø Godøy)

1.1. Opening of the Session

The meeting was opened on Sept 14, 2017 at 9:00 by Dr Barry Goodison, Vice-Chair of the Global Cryosphere Watch (GCW) Steering Group (GSG) who welcomed the participants and extended best wishes for a successful meeting, highlighting the importance of the progress made in the development of the GCW Data Portal, and appreciating the significant contribution of the Norwegian Meteorological Institute (MetNo), which is hosting the GCW Data Portal.

On behalf of the Secretary-General of the World Meteorological Organization (WMO), Prof Petteri Taalas, Ms. Rodica Nitu also welcomed the participants, noting that the WMO activities focusing on the Polar and High Mountain Regions are one of the seven priorities for WMO and its Members during the current financial period. This meeting is an important milestone in the development of the GCW as an operational programme of WMO, as decided by World Meteorological Congress, Cg-17 (2015).

1.2. Purpose of the meeting

Dr Øystein Godøy, the Chair of the GCW Data Portal Team noted the importance of the meeting towards an operational GCW, inviting the participants to contribute to the following objectives:

- Review the GCW Operational Manual for contributing data centres (referring to System management) and GCW Guide for data centres contributing to GCW (referring to the technical implementation);
- Develop a road map of engagement for the development and testing of the GCW Interoperability Package for CryoNet stations;
- Discuss requirements for compliance to WIS and WIGOS;
- Develop a road map for advancing the development of standard vocabularies for metadata for the cryospheric variables recommended for the observing programs of GCW (CryoNet).

1.3. Adoption of the Agenda

The meeting agenda was adopted as presented, and is available in **Annex 1**, of this report.

1.4. Working Arrangements for the Session

The List of Participants is attached in **annex 2.**

The group reviewed the security arrangements for the venue, and agreed on the working schedule for the two-days meeting.

It was noted that:

- Dr. Julie Friddell would join in the session by WebEx;
- Dr. Peter Pulsifer and Dr Maribeth Murray will attend only the day two of the meeting;
- Prof Shawn Marshall will attend only day one of the meeting.

2. Status of development of interoperability with other Data Centres

Dr Øystein Godøy provided an overview of the current status of the development of the GCW Data Portal and its connectivity with data centres (http://globalcryospherewatch.org/data/data.html). As noted, the GCW Data Portal is hosted by MetNo (http://gcw.met.no/metamod/search .

He indicated that the GCW Data Portal is a mechanism to provide access to relevant cryosphere datasets, real time data streams, as well as access to archived data. MetNo does not host data,

although there could be a possibility for doing so, at some point in the future. The interoperability is critical to data sharing and to the success of GCW as an operational programme.

Ø Godøy noted that the Data Portal is developed as a Distributed Data Management system, being metadata driven, net centric (linkages with other data centres are vital, and implies the brokering of data and metadata), and interdisciplinary (dataset agnostic in the open data space). It's expected to be a WMO Information System (WIS) Data Collection and Production Centre (DCPC), and to be compliant to the WMO Integrated Global Observing System (WIGOS) regulatory framework. So far the primary focus has been on WIS integration.

- The discovery metadata are generated by the data centres hosting the data sets, and are harvested at regular intervals by the Data Portal using the Open Archives Initiatives Protocol for Metadata Harvesting (OAI-PMH[1]) and the Open GeoSpatial Consortium – Catalog Services for the Web (OGC CSW[2]). Both these are currently under test, and are ingested in the central catalogue for usage by the GCW Portal user community.
- Currently, the portal does not utilize a distributed search, although it may do so in the future.
- The Data Management services of the Data Portal use OPeNDAP[3] (Advanced Software for Remote Data Retrieval) for integration of data.

It was mentioned that the Portal links with a heterogeneous community of data centres (research and operational), with various degrees of interoperability for metadata, data, standardization, support, programing languages. GCW Data Portal has been used as the backbone for the data portal of the Year of Polar Prediction (YOPP), for the Svalbard Integrated Arctic Earth Observing System (SIOS) https://sios-svalbard.org/, and for the Data Portal of the Arctic Polar Regional Climate Centre (Arctic PRCC).

Ø Godøy reviewed some of the challenges in developing the Data Portal, which need to be taken into account:

- One is cultural; while the operational communities are driven by compliance, and are more hierarchical, the research communities are driven by pragmatism, operating often as informal networks, based on guidelines (more than one way to accomplish results).
- The other challenge is related to the implementation in practice of interoperability.
    - For discovery metadata, protocols and structures are available; however the availability and the consistency of semantics and terminology remain a significant gap. For the discovery of metadata and data, controlled vocabularies with definitions well accepted by the scientific community, are needed.
    - While protocols exist for data exchange, there are major gaps in consistent formats, semantics/terminology as well as common data models (e.g. UNIDATA Common Data Model, CDM) are major gaps. Standards for these needs to be developed and adopted at large scale to facilitate the exchange of information. Legacy systems do not have to be changed, as long as they could be mapped in the common data model, which is a major priority.

As a WMO activity, GCW Data Portal is committed to WIS, and the application of the ISO19115 WMO Profile (WIS metadata) for the discovery of data. Concerning description of the observation facilities,

1 https://www.openarchives.org/pmh/
2; http://www.opengeospatial.org/standards/cat;
3 https://www.opendap.org/

the WIGOS metadata standard (based on OGC Observations and Measurements) is currently under development, in particular regarding the cryosphere variables. The WIGOS metadata describes, inter alia, the measured variables, as well as the temporal evolution of instrumentation, procedures of measurement, processing, local conditions, etc., required to fully understand the data found.
It was noted that the Data Portal includes data transformation functions which, once the data are available, would allow users to extract data and generate and compare products tailored for their specific needs, e.g. data series, mapping, visualization, etc.
The Data Portal is supported on the MetNo hardware infrastructure, which is based on a private cloud and a scalable high performance computing infrastructure.

The group recalled that each CryoNet station will receive its own WIGOS ID number. This is allocated at national level by the Permanent Representative (PR) of the respective country. Alternatively, the WMO Secretary General could assign one, although this is not the preferred approach.
Action: GCW PM to remind GCW Focal Points about their role in facilitating the issuance of WIGOS IDs, at national level. Also, reminding the CryoNet and contributing stations representatives the need to obtain PR endorsements for their respective stations.

If data is needed in real time, MetNo could collect data convert it to BUFR and put it on GTS, or this could be done via the national meteorological service, when/if the users ask for it. The bandwidth is an issue in some parts of the world. GTS was set up to support real time exchange of the necessary information for WMO purposes.

3. Review of GCW Interoperability Guidelines

Ø Godøy noted that two documents are under development and will guide the implementation of the GCW Data Portal. These are:

- GCW Operational Manual for contributing data centres (referring to System management).
- GCW Guidance for data centres contributing to GCW (referring to the technical implementation), and

The group recalled that the CryoNet stations have committed to share their data, according to the GCW recommended practices, and this includes discovery metadata. The Data Portal functionality expects that stations in the GCW network adopt discovery metadata at the lowest granularity, to enable discovery of data and advanced analysis. Currently, the application of protocols and the understanding of interoperability are highly variable among the contributing data centres. As is the granularity (e.g. describing individual datasets or collections of datasets) of the discovery metadata provided.
To control costs and involved resources, and to achieve an effective system, the number of interface standards used for metadata exchange needs to be limited, and reuse external services where possible.

3.1 Review of GCW Operations Manual for contributing Data Centres

Ø Godøy presented an overview of the GCW Operations Manual for Contributing Data Centres, currently under development. He invited all participants to review the draft manual and provide feedback, reflecting their specific circumstances**.**
This manual describes the procedures necessary to maintain an operational GCW Data Management system over time, and its target audiences are the system managers maintaining the data

management systems at the data centres contributing to the GCW Portal. A structured change management system is included and is critical to facilitating the availability of up to date information on changes of end point, thus avoiding breaks in communication.

It was highlighted that the contributing data centres are responsible for (1) setting up metadata endpoints exposing metadata according to the GCW guidelines, (2) informing the GCW Data Portal of interruptions in the availability of endpoints, (3) are responsible for the quality of the metadata made available, and (4) for applying a Request for Change (RFC) change management system, when changes to the local system are required. This will avoid interruption and conflicts in metadata harvesting, as outlined in the Manual.

Discussion points:

- Many CryoNet stations don't have data management systems, and GCW needs to help them. More considerations are needed to ensure that GCW is effective as a service.
- A working Change Management System is critical for a functional distributed system. Endpoint changes need to be communicated and monitored. The group supported the proposal for a Request for Change system to track changes to the system, potentially using a BitBucket to control access to individual projects.
- Key Performance Indicators (KPIs) are needed to ensure system monitoring and performance improvement. In a distributed system, the bandwidth is an issue, and could impact monitoring.
- M Murray emphasized the value of accurate KPIs in support of the preparation of funding proposals.
- KPIs proposed, are: the number of data sets in a catalog, the number of visitors per catalog, the use of DOIs for tracking how the data is being used. Alternative metrics (how many mentions on Twitter, Facebook, etc…. as a measure of impact). More to be added.

Actions:

- **Review and update the provisions of the Operations Manual after the completion of the pilot project with SLF (see Section 4 of this report).**
- **Define what is expected from a CryoNet station, regarding the availability of CryoNet station data, e.g. whether WMO requires the number of datasets exchanged.**
- **GCW to provide support to CryoNet and contributing stations regarding the adoption of standard metadata (WIGOS, OSCAR). Recommendations to be made where to link to WIGOS metadata, to OSCAR metadata or to a local copy.**
- **T Kralidis and Ø Godøy to develop a wiki with guidance, tips on the implementation of metadata (using Trello or GitHub).**

It was noted that, for those CryoNet stations which are not from an NMHS, adopting the WIGOS metadata is a laborious process, if needed to be entered manually. The GCW Data Portal could offer a service in expediting the entry of metadata from the small data centre.

- **Action: Ø Godøy to share details on how the Data Portal could help small data centres regarding the entry of WIGOS metadata, and what else is needed to be developed/made available for it.**

3.2 Review of the GCW Guidance for data centres contributing to GCW

Ø Godøy noted that the main goal is accessing the data, extracting the information, which includes generating BUFR for exchange data in real time, however adopting standard metadata by each data

centre, is critical to discovering the available data. The guide defines the following levels of metadata:

- Discovery metadata for identifying relevant data products; this is implemented, held, and provided by the data centres in GCMD DIF[4] or ISO 19115 (i.e. WIS metadata) formats.
- Configuration metadata, used for tuning for specific datasets and services, e.g. visualization or transformation; this is created by the GCW Data Portal using information harvested from data centres.
- Use metadata, for understanding the data accessed, e.g. NetCDF files formatted according to the Climate and Forecast Convention.
- Site metadata, for understanding the context of the observations, which is linked to WIGOS metadata (station location, surroundings, instrumentation, procedures).

GCW needs to be pragmatic, as most data centres contributing to the GCW Data Portal are not linked to the national meteorological services. The success of interoperability between the Data Portal and each CryoNet or contributing station depends on the support that GCW could provide for enabling the understanding of WIGOS and WIS framework.

Ø Godøy highlighted some of the key points included in the Guide. Specifically, regardless of the metadata standard used and the mechanism for transport of the information, the Data Portal recommends that the contributing repositories ensure that all datasets have a unique identifier for tracking datasets in the central repository and check for duplicates. The identifier is set by the authoritative source for the dataset. The GCW Data Portal will not specify or change a unique identifier unless the dataset is hosted by the GCW Portal; this is not supported, right now.

The OAI-PMH protocol is the recommended interface for exchanging metadata with the GCW Portal. It is a cost effective, robust mechanism for exchanging metadata between data centres, used by WMO WIS and is under consideration for WIGOS metadata exchange. A number of software solutions supporting it, are freely available. All data centers planning to become interoperable with GCW Data Portal are requested to use OAI-PMH version 2.

Ø Godøy strongly recommended that the large repositories containing much more than GCW relevant data, to configure dedicated cryosphere or GCW sets (action, all Data Centre Managers), to reduce the filtering of harvested metadata performed by the Data Portal. The name of the set that GCW should harvest has to be communicated to the Data Portal manager, and names like “GCW” or “Cryosphere” are recommended. More information is available in OAI-PMH Set specification.

- In the current configuration, GCW harvests from partner nodes in a manner where the GCW node provides a specific 'set' of discovery metadata pertinent to GCW only via the OAI-PMH standard. This allows a client to specify a 'set' parameter for broad filtering.
- T Kralidis recommended that the GCW node provides an OGC CSW server, for harvesting with filters. Ø Godøy noted that this approach will be supported in the future.

Ø Godøy requested all Data Centre Managers, as a matter of practice, to notify the Data Portal Managers when records are deleted in the contributing data centres catalogues (action) (see OAI-PMH specification of deleted records).

4 GCMD DIF: Global Change Master Directory - Directory Interchange Format, (https://gcmd.nasa.gov/)

Metadata has to be provided to the GCW Data Portal in ISO19115 and/or GCMD DIF (action, Data Centre managers). The OAI-PMH interface offers by default metadata in Dublin Core, which is not sufficient for GCW purposes. In order to properly identify the metadata standards in the responses provided by the OAI-PMH end point, it is recommended to use the following keywords: "dif" for GCMD DIF, "iso" for ISO19115 minimum profile, "wis" for the WMO Core Profile of ISO19115 and "wigos" for WIGOS metadata in the "ListMetadataFormats" response. The latter is yet not fully defined in XML.

The Open Geospatial Consortium Catalogue Services for the Web (OGC CSW) is another standard for exposing the content of a catalogue in a standardized form, alternative to the OAI-PMH. If used, GCW recommends that OGC CSW version 2.0.2 is used. As multiple data streams are expected to be going out to different centres, the GCW Data Portal must ensure that these are consistent, and avoid duplicates and minor differences.

Currently there are no controlled vocabularies for most of the cryosphere variables. GCW has initiated the development of a GCW set of key words and their associated definitions, and GCW needs to encourage Data Centres to use the recommended GCW keywords (action, Secretariat, C Hrynkiw).

Ø Godøy noted the activities planned on the Data Portal, in the next period:

- Implementation of the user interfaces (currently under test), shopping cart to download data as a web service (under test), and the metadata interfaces;
- Development of interfaces for WIGOS metadata;
- Promote the implementation of DOIs for traceability on data usage, as a mean to provide recognition to the values of data.
- Single sign on solutions: currently on hold, but prepared for SAML[5] and OAUTH2[6].

B Goodison noted the significant differences between the capabilities and engagements of the participating Data Centres, and requested an update on engagement of the Data Portal engagement with centres like the World Glacier Monitoring Service (WGMS[7]). Ø Godøy noted that the funding model of WGMS requires that they promote, primarily, their services. The interoperability with WGMS is on hold, right now, as there is no support for documenting the data, only the service. This works for WIS, but not for GCW, as the goal is to integrate data from neighboring, but independently operated, stations. **B Goodison requested that the team and Secretariat follow up with WGMS (action).** S Marshall noted that the Kluane Lake site will link to WGMS, as well. This is similar to Norway where the glacier stations are included in WGMS and Cryoclim (http://cryoclim.met.no/). There are glacier parameters which are not of interest to WGMS, but could be made available via GCW, for other applications.

4. Introduction GCW Interoperability Package for CryoNet stations

J Fiddes presented an overview of the GCW Interoperability Package for CryoNet stations, which is an initiative of GCW, and Dr Joel Fiddes has been hired by WMO for this task (Annex 3).

5 SAML: Security Assertion Markup Language

6 OAuth 2 is the industry-standard protocol for authorization ((https://oauth.net/2/)

7 http://wgms.ch/

This is the second phase of a demonstration project for coupling a data centre of a CryoNet station with the GCW Data Portal. This builds on the success of coupling a research data-centre of the Weissfluhjoch CryoNet site with the GCW Data Portal, which was initiated in 2015 in collaboration with the Swiss Federal Institute for Snow and Avalanche Research (SLF). In this second phase, the focus will be on working with a small data centre. The work with SLF demonstrated the dynamic sharing of metadata using the OAI-PMH protocol, and has become a template for interoperability with other data centres contributing to the GCW observing network. The proof of concept of the interoperability package is expected to be completed by April 2018.

The planned application will be a "sit–on-top" solution, for easy integration into existing data-management practices, using standard protocols (i.e. OPeNDAP), preferably a GUI based metadata editor + server (OAI-PMH), and utilize as much as possible existing protocols and applications, and keep costs to a minimum.
This development will be coordinated with further development at SLF. The system includes the data processor, a webserver, a data discovery, as well as possibly a visualization component.

- The Data Processor contain the MeteoIO application, which is an integral part of SLF avalanche warning service, and benefits from long-term support (https://models.slf.ch/). It reads different data types from various sources and provides a standard output, using a plug-in architecture.
- The Webserver operates with OPeNDAP, and allows the streaming of data recombined in virtual datasets, without any file transfer.
- The data discovery EnviDat is SLF's flagship data portal project, based on CKAN (used in Sweden, Canada, US, UK), and it benefits from long term support by SLF, http://www.envidat.ch/dataset/asrb-wfj .
- For visualization, two existing applications are available, http://www.osper.ch/ and http://snopviz.org/

The participants agreed that or the additional development, J Fiddes will work with a newly established CryoNet site (Kluane Lake, Canada), a well-established DCPC (ECCC), and another well-established CryoNet site (Sodankylä); action. Similar collaboration is desired with the CryoNet stations in China. The existing relationship with SIOS, at MetNo, makes it reasonable to consider this integration, as well.
The completion of the package depends on the additional development at SLF. Close coordination with SLF will help in addressing any potential issues as soon as they arise.

It was noted that the data logger manufacturers are interested in receiving recommendations on standard output formats to ensure compatibility, e.g. NetCDF.
Action: GCW works with CIMO and HMEI[8] on promoting standard formats for data logger outputs which could facilitate the standardization of data formats for cryosphere variables.

Next steps (actions);

- **J Fiddes will coordinate with SLF the development of the concept.**
- **GCW will conduct a survey of all CryoNet stations to determine the current configuration of their systems and data management methods, in order to identify the**

8 HMEI: Hydro-Meteorological Equipment Industry Association,

user requirements and to ensure their engagement in the development of this proof of concept.

- **For engagement a forum to collect info, will be setup. Trello will be used for collaboration. The experience and results of the work planned under the current engagement will constitute the basis for further evolution, as identified during the current development process. J Fiddes to create a space on Trello to invite all interested parties to contribute to this development.**
  For the development and demonstration of the package and the CryoNet survey, J Fiddes will collaborate with S Marshall for Kluane Lake CryoNet station, T Kralidis on behalf of ECCC, and K Luojus (Finnish Meteorological Institute, for the Sodankylä CryoNet site).

Discussion points:

- The group agreed that interoperability in terms of format standards are needed, and standard vocabularies require the engagement of scientific and data communities.
- OpenSearch is an alternate application for harvesting metadata, and is currently not used by GCW. The planned software is modular and could be used in parts, as needed.
- Service interfaces are needed for the integration of applications developed in different programming languages; e.g. EnviDat: is based on Python and MeteoIO in C++. MetNo is working with gridded data in C++ with a Python shell, to enable the link with the Python community.
- The proposed package addresses the GCW needs. Many GCW stations are linked to other programs, supporting other applications, and the coordination between these applications and programs, internationally, is increasingly, of interest for the international community.
- While ecology is a system science, before proceeding with system standardization, we needed to get the underlying disciplines standardized, first. The goal is to build portals which integrate (machine to machine) with other portals to enable discoverability at portal level: "*one submission, one data record*" approach, available to different platforms.
- Funding agencies are increasingly mandating the sharing of data, and data citation guidelines are needed. Generating data is a real contribution to scientific goals, and each station contributor needs to attach s DOI obtained nationally.
- The INTERACT network is a potential beneficiary of this package. Kluane Lake is a member of INTERACT, and other GCW stations are, as well. At the INTERACT meeting in October 2017, Ø Godøy will promote this application and develop broader engagement.
- The value of being part of CryoNet is that a station has higher visibility as part of a bigger network, which improves the chances for funding; e.g. Kluane Lake CryoNet station has an opportunity for leveraging support from the Yukon Government.
- It was agreed to postpone the implementation of data quality filters to a later stage. The participants agreed that these filters should be included in the Best Practices Guide. Of all, the flagging of missing data is a filter important to the users of data.

Action: Data Portal Team to ensure that the Best Practices Guide reflects guidance on data quality

- The hydrological data are equally relevant to cryosphere studies and the availability of metadata and data via the GCW Data Portal is highly desirable.

Action: GCW Observations WG to assess how to include the hydrological data, in addition to meteorological data, as ancillary information.

- The GCW Focal Points need to be more actively engaged in facilitating the dissemination of information regarding the implementation of interoperability and assessing how they can influence the development at national level.

Action: Secretariat to communicate regularly with the Focal Points, and ensure that the terms of reference reflect their expected active engagement, including on the data and metadata exchange.

5. Proposed Polar Data Forum

P Pulsifer, the Chair of the Arctic Data Committee (ADC[9]) of SAON[10] introduced the concept of a potential polar data forum, to take place in 2018, with (partial) funding from the National Science Foundation (NSF). The goal of the forum is to develop a statement for the Arctic Science Ministerial (2018), by promoting data standardization and interoperability at system level, and by engaging the scientific and operational communities active in the arctic. He noted that there are many organizations collect observations in the arctic, at different scales, that there is a high cost for observations in the arctic and, also, high pressure. The value of observations increases through integration and there is a need to sensitize policy makers of the role and cost of data necessary to sustain all other components of arctic engagements. **P Pulsifer invited GCW to be an active contributor in the organization of the forum, recognizing the work underway on the interoperability of the Data Portal and the other data centres, and the GCW representatives agreed to be engaged. (action)**

This topic was also discussed at the ADC meeting on Sept 16-17, 2017, where GCW was also represented.

The work of GCW regarding the development of standard vocabularies is very relevant for the entire arctic community. The forum would bring together key players like GCW, INTAROS[11], INTERACT[12], SAON/ADC, GEO[13] and align the community needs and contributions. The Arctic Council Chairmanship by Finland is an opportunity to facilitate the increased focus on data needs. Recently, the representative of Finland on SAON proposed the project "Arctic Observation Value project - assessment for physical atmospheric and oceanic variables", which is an appropriate framework for the proposed focum.

6. Overview of the CryoNet stations in China

Dr Feiteng Wang (State Key Laboratory of Cryospheric Science of the Chinese Academy of Science, CAS) provided an overview of the Cryospheric observation network of China, highlighting the contribution to and engagement with GCW.

Five of the cryosphere stations operated by the State Key Laboratory of Cryospheric Science of CAS, are also approved as CryNet sites or stations; these are:

- Tienshan station (member of WGMS),
- Germu station (permafrost research and monitoring along Quighai-Tibetan Plateau highway and Railway),

9 ADC: Arctic Data Committee
10 SAON: Arctic Sustained Observing Networks
11 INTAROS project, working towards an integrated arctic observing system
12 INTERACT: International Network for Terrestrial Research and Monitoring in the Arctic
13 GEO: Group on Earth Observations

- Qilian station (focusing on observations on glacier, permafrost, snow cover, atmosphere, hydrology, ecology in the Shule river basin),
- Tanggula Station, focusing on monitoring cryospheric components in the Dongkemadi River basin in the inland Tibetan Plateau , and
- Koxkar station, located on Mt. Tuomuer, focusing on observation in Tuomuer-type glacier as debris covered valley glacier.

Other stations which are part of the same CAS program, are Yulong Station (founded in 2006, Yunnan province, a field observation platform for China's monsoonal temperate glacier in low-latitude region), Altai Station (founded in 2017, located at Jimunai, Xinijang, providing cryosphere science and technological services for the sustainable development in the region), and the recently established station in Barrow (Alaska) as a joint project with University of Alaska.

Other stations operated in Polar Regions are the Yellow river station, Kunlum station, GreatWall station, and Zhongshan station.

F Wang noted that The State Key Laboratory of Cryospheric Science of CAS has a well-developed cryospheric sciences data platform, the Cryosphere Science Data Platform, http://csd.casnw.net , which is, currently, only available in Chinese. Plans are being developed to develop an English translation and participate in the integration of data with GCW.

Action: J Fiddes to involve CAS in the development of the application for data exchange pilot.

Ø Godøy noted that the GCW Data Portal is already harvesting metadata from the Polar Research Institute of China. GCW needs to better understand the linkages between CAS and the Polar Research Institute of China, and how the latter can facilitate the development of interoperability between CAS stations and the GCW Data Portal.

Action: GCW Secretariat to provide additional information regarding the linkages between CAS, CMA and the Polar Research Institute of China, and how these could help in improving the interoperability with the Data Portal.

7. Development of Standard Vocabulary and Terminology

The group agreed that the development of controlled vocabularies to complement the existing WIGOS/OSCAR metadata, and reflect the variables proposed for CryoNet stations, is one of the highest priorities for the upcoming period, for GCW.

Currently, the WIGOS Metadata Representation, for the WMO Integrated Global Observing System (WIGOS), WMO-No. 1160, is currently under development. Some cryospheric variables are already included, some of those have definitions, but for most of them definitions need to be added, reflecting a broader acceptance by the scientific community, which is critical to their use. A crosscheck with CF discovery metadata and station metadata is needed.

Charmaine Hrynkiw (ECCC) will be supporting GCW in identifying and addressing the gaps on the availability and definition of cryosphere variables, and the associated metadata, leading to a proposal for controlled vocabulary for cryosphere data, using the list of recommended variables, ((http://globalcryospherewatch.org/cryonet/variables/recommended_variables.html), and the GCW Glossary (http://globalcryospherewatch.org/reference/glossary.php) focusing on what is reasonable and making proposals for addressing the differences.

The goal is to map WIGOS and OSCAR variables with what is available in CF. The WIGOS Metadata task Team will meet on Nov 27-29, 2017 and input from GCW on the cryosphere variables is essential, if approval is to be received for inclusion in the WIGOS manual at the EC-70 (June 2018).

The group reiterated the need for regular review and updates. It was recommended to establish a web environment, with broader access, for peer review of the proposed terminology.

Actions:

- **C Hrynkiw will focus first on those variables which are exchanged or would be exchanged on GTS (snow depth, SWE, etc.). Use Trello as a consultation board.**
- **Identify variables by categories (well-known and clear, uncertain, missing definitions, variables not included).**

It was recommended to take what we have and develop a demo project to translate it into a machine readable form (ISO code list), and seek the community agreement, e.g. use short and extensive definitions, and taking into account the context of the definition (obs, modeling, etc.)

The group noted as a valuable resource, the Glossary from Research Data Canada (https://www.rdc-drc.ca/glossary/), the CASRAI[14] Dictionary for Research Data Management, a comprehensive list of research data management terminology, and including polar terms, as well.

Ø Godøy noted that currently, there are calls in Europe for projects on terminology and semantics and NSF has similar research coordination activities. He reiterated that GCW will align its efforts with those of other initiatives with similar. **It was agreed that GCW needs to remain informed and engaged with related projects from other initiatives, like those in Europe.** For example, GEO Data Management initiative approach of "crawling the web for terms" could contribute to ensuring that nothing is missed, also, the ADC has a subgroup which is focusing on consolidating vocabularies and have already reviewed the GCW Glossary terminology (e.g. ADC[15]-IARPC[16] Vocabulary and Semantic Working Group).

8. General Discussions

The group received the proposal for developing the GCW Interoperability package for GCW (Section 4 of this report) very positively, and recommended that **communities of potential users are notified about the development of this software, and are invited to provide input and suggestions** (action). Once the interoperability package is available, the value Data Portal could be demonstrated with the products possible to be derived with the data made available. M Murray noted that the contribution of Kluane Lake is expected to be of value to the community, as it plans on getting 1-2 stations on line, in the near future and this will be a good demonstration of the viability of the proposal.

M Murray emphasized the critical value of a sustainable system, and the need for long term commitment.

P Pulsifer recommended getting traction through different communities**.** Use Twitter to spread the word and demonstrate to people that the code is reliable and sustainable, and the community around it will keep it going. There is a need to engage more people by talking to the communities which are active in arctic research, to ensure that the code could be made sustainable through the broader contribution of experts. He strongly suggested identifying a small number of projects to collaborate on and sustain and promote. Identify teams to work on different components: dev, packaging, documentation, early adopters.

14 CASRAI: Consortia Advancing Standards in Research Administration
15 ADC: Arctic Data Committee of the Arctic Sustained Observing Networks (SAON)
16 IARPC: Interagency Arctic Research Policy Committee

It was recommended that J Fiddes develops the proposed application for data access and engages others to develop other components. With a broader engagement it will be easier to demonstrate the value added.

M Murray encouraged GCW to **develop a strong documentation base which is paramount for smaller facilities which, often, rely on one expert for all functions (Action, GCW)**. This would support the development of capacity and ensuring sustainability, while reduce the vulnerability of depending on one expert. The documentation should include operations and maintenance guides/manuals, as well as a community space to post testimonials; deployment documentation is critical to help with the decision on joining, developing, sustaining a station; it was suggested to create virtual machines to DOCKER (virtual machines).
P Pulsifer shared that all NSF new contracts for partnering include a requirement for virtualized technology to be used and be cloud accessible.

The group discussed also:

- The need to effectively bridge the gap in understanding between science and data management in support of science. The proposed interoperability package has value for data managers and for scientists, equally, but this needs to be demonstrated and seek to have it adopted.
- The elements of a data stream, and whether the processing of data is part of the metadata, recognizing that increased complexity of measuring technology requires increased processing to derive output data products. P Pulsifer recommended that the visualization of uncertainty and error are indications of value, at discovery level and at the data level, and provide feedback to scientists, and users based on the assessment of data and metadata.

The group agreed about the need to be able to share and index images, sound, videos, as information and tools are, increasingly, accessible. The key is to engage end users, to ensure that they have an interest, and define how to manage and share information, in general.
The images could be addressable and could be combined with the disc metadata. More automated image processing requires additional considerations for protection, sustain, power, include remote sensing.
Action: GCW Data Portal to take into account the need for sharing, exchanging information as images, videos, etc.

9. Next Steps

Other actions:

- CryoNet sites to provide to J Fiddes test cases and try with csv files, testing based on functionality in the system;
- Secretariat to work with C Hrynkiw to align the schedule of the Metadata work with the schedule of the Metadata TT meeting (Nov 27-29, 2017) and deliverables.
- Organize an informal follow-up at the Arctic Change Assembly 2017 (Quebec City). A presentation on GCW to be submitted for the conference, and to reference the work on Best Practice Guide development.

- J Fiddes, Ø Godøy, and S Marshall to collaborate on the development of applications for data discovery and exchange.
- B Goodison recommended that GCW leads the organization of a side event EC-70 on the exchange of info (snow depth, SWE).

10. CLOSURE OF THE SESSION (Ø Godøy)

The session was closed on Friday, September 15, at 17:00.

Annex 1: Meeting Agenda

1. ORGANIZATION OF THE SESSION
- 1.1 Opening of the Session
- 1.2 Adoption of the Agenda
- 1.3 Working Arrangements for the Session

2. Outline of GCW data management including relations to the WMO Information System - WIS - and the WMO Integrated Global Observing System – WIGOS -

3. Status of development of interoperability with other Data Centres

Overview of the current status of the development of the portal and its connectivity with data centres

4. Review of GCW Interoperability Guidelines
- 4.1 Discovery metadata
- 4.2 Use metadata
- 4.3 Exchange of discovery metadata
- 4.4 Exchange of data

5. Introduction GCW Interoperability Package for CryoNet stations
- 5.1 Outline
- 5.2 Concepts
- 5.3 Deliverables
- 5.4 User engagement

6. Review of GCW Operations Manual

7. Future work
- 7.1 User engagement (questionnaire)
- 7.2 SLF integration
- 7.3 Canadian integration?

8. OTHER BUSINESS

9. CLOSURE OF THE SESSION

Annex 2: List of Participants

	Participant	Affiliation	e-mail
1	Barry Goodison	Vice Chair, GCW Steering Group Canada	barrygo@rogers.com
2	Øystein Godøy	Norwegian Meteorological Institute Chair, GCW Data Portal Team	o.godoy@met.no
3	Eivind Støylen	Norwegian Meteorological Institute Lead, of Arctic PRCC, Norwegian Node	eivinds@met.no
4	Kari Luojus	Finnish Meteorological Institute Co-Lead, Snow Watch Team Co-Chair, GCW Integrated products WG	kari.luojus@fmi.fi
5	Peter Pulsifer	Chair, Arctic Data Committee, University of Colorado	Peter.Pulsifer@Colorado.EDU
6	Tom Kralidis	Environment and Climate Change Canada	Tom.kralidis@canada.ca
7	Ross Brown	Environment and Climate Change Canada Co-Lead, Snow Watch Team	Brown.Ross@ouranos.ca
8	Anne Walker	Environment and Climate Change Canada GCW Focal point for Canada	anne.walker@canada.ca
9	Maribeth Murray	Arctic Institute of North America University of Calgary	murraym@ucalgary.ca
10	Shawn Marshall	University of Calgary Representative, Kluane Lake CryoNet station	marshals@ucalgary.ca
11	Feiteng Wang	Cold and Arid Regions Environmental and Engineering Research Institute (CAREERI), Chinese Academy of Science (CAS) GCW Observation WG Representative of CAREERI CryoNet stations (China)	wangfeiteng@lzb.ac.cn
12	Julie Friddell	Canadian Cryosphere Information Network, University of Waterloo GCW Data Portal Team	julie.friddell@uwaterloo.ca
13	Eleanor Blackburn	Environment and Climate Change Canada Point of Contact for CryoNet stations (ECCC)	Eleanor.blackburn@canada.ca
14	Charmaine Hrynkiw	Environment and Climate Change Canada	charmaine.hrynkiw@canada.ca
15	Joel Fiddes	WMO Consultant	joelfiddes@gmail.com
16	Rodica Nitu	WMO Secretariat	rnitu@wmo.int

Annex 3: GCW INTEROPERABILITY Package

GCW Open Source Software (OSS) data exchange, supporting the interoperability of the GCW Data Portal with other data centres

Terms of Reference

MOTIVATION

One of the most important milestones in achieving full interoperability between the GCW Data Portal and the individual CryoNet stations data centres is the ability to dynamically exchange metadata and data.

The Resolution 43 of Cg-17 decided to implement the Global Cryosphere Watch (GCW) in the WMO Programmes as a cross-cutting activity, during the 17th financial period as one of the major efforts of the Organization; GCW should become operational as a programme of WMO, by 2020. Decision 50 of EC-68 acknowledged the need for speedier operationalization of the GCW Data Portal, hosted by Norway, by expanding its interoperability with major data centres and all CryoNet stations/sites. Decision 45 of EC-69 approved 120 stations as CryoNet and Contributing stations of the GCW surface observing network; more than 50% of these are operated by universities and research organizations, which are not familiar with the WMO metadata and data exchange mechanisms and do not have the resources to adopt them. Yet, the exchange of data and metadata is one of the most important benefits that GCW offers, and the interoperability of the GCW Data Portal must provide a software solution to address this.

A demonstration project for coupling a small data-centre with the GCW Data Portal was initiated in 2015, in collaboration with the Swiss Federal Institute for Forest, Snow and Landscape Research (SLF). This demonstration project was successful in the dynamic sharing of metadata using the OAI-PMH protocol. The results achieved have provided a template for interoperability with data centres contributing to the GCW observing network.

This proposal focuses on follow up work from the demonstration project with SLF, and it primarily aims to consolidate the results achieved with SLF and develop a design specification for a data-system that is open-source, easy to deploy, satisfies GCW requirements (and further contributes to defining those requirements) that is targeted at small/medium sized data-centres. The results expected would build on ongoing developments at partner institutions such as the SLF/WSL.

Deliverables

1. Report on the status of the data management systems operated by each organization participating with stations to the Global Cryosphere Watch and their readiness to exchange data and metadata with the GCW Data Portal (including by surveying the participating organizations): Due 31.12.2017
2. Further develop the GCW Interoperability Guidelines with emphasis on data interoperability using NetCDF and OPeNDAP. Includes descriptions on how to utilize OPeNDAP to retrieve the real time data directly from the OPeNDAP stream: Due: 10.01.2018
3. Develop proof of concept for using existing SLF interfaces for other data-centres (software and methodology), showing how to use the developed open-source applications to deliver a GCW compatible data-system. Test the developed concept with a data centre of an at least one other CryoNet station: Due 31.03.2018

4. Prepare a report identifying the additional development needed (e.g. Promoting CF compliant metadata) for full interoperability between the GCW Data Portal and each of the Data Centres contributing to the GCW observing network: Due 31.03.2018.

Annex 4: Action Items

No.	Reference	Action	Deadline	Responsible
1	2.0	Action: GCW PM to remind GCW Focal Points about their role in facilitating the issuance of WIGOS IDs, at national level. Also, reminding the CryoNet and contributing stations representatives the need to obtain PR endorsements for their respective stations.	GSG-5	GCW PM
2	2.0	Review and update the provisions of the Operations Manual after the completion of the pilot project with SLF (see Section 4 of this report).	2018	Ø Godøy
3	2.0	Define what is expected from a CryoNet station, regarding the availability of CryoNet station data, e.g. whether WMO requires the number of datasets exchanged.	2018	Ø Godøy; W Schöner
4	2.0	GCW to provide support to CryoNet and contributing stations regarding the adoption of standard metadata (WIGOS, OSCAR). Recommendations to be made where to link to WIGOS metadata, to OSCAR metadata or to a local copy.	2018-19	GCW
5	2.0	T Kralidis and Ø Godøy to develop a wiki with guidance, tips on the implementation of metadata (using Trello or GitHub).	2018	T Kralidis and Ø Godøy
6	2.0	Ø Godøy to share details on how the Data Portal could help small data centres regarding the entry of WIGOS metadata, and what else is needed to be developed/made available for it	2018	Ø Godøy
7	3.2	Ø Godøy strongly recommended that the large repositories containing much more than GCW relevant data, to configure dedicated cryosphere or GCW sets	On going	CryoNet/contributing station managers
8	3.2	Ø Godøy requested all Data Centre Managers, as a matter of practice, to notify the Data Portal Managers when records are deleted in the contributing data centres catalogues (action)	On going	CryoNet/contributing station managers
9	3.2	Metadata has to be provided to the GCW Data Portal in ISO19115 and/or GCMD DIF according to the instructions provided in the GCW Guidance for data centres contributing to GCW (action, Data Centre managers).	On going	CryoNet/contributing station managers
10	3.2	GCW to develop a GCW set of key words and their associated definitions, and GCW needs to encourage Data Centres to use the recommended GCW keywords.	2018	Ø Godøy

	3.2	Data Portal interoperability with WGMS is further explored	GSG-5	GCW PM
11	3.3	J Fiddes will work with a newly established CryoNet site (Kluane Lake, Canada), a well-established DCPC (ECCC), and another well-established CryoNet site (Sodankylä), for advancing the data interoperability concept of GCW.	2018	J Fiddes; S Marshall; T Kralidis; K Luojus
12	3.3	GCW works with CIMO and HMEI on promoting standard formats for data logger outputs which could facilitate the standardization of data formats for cryosphere variables.	2019	GCW PM (to facilitate)
13	3.3	J Fiddes will coordinate with SLF the development of the concept of the interoperability package	2018	J Fiddes
14	3.3 & 7	GCW will conduct a survey of all CryoNet stations to determine the current configuration of their systems and data management, to identify the user requirements and to ensure their engagement in the development of this proof of concept. It will be used as a mean to inform Station Managers and Focal Points about the planned activities.	11.2017	J Fiddes; GCW PM
15	3.3	Trello will be used for collaboration with all engaged in the interoperability pilot project. J Fiddes to create a space on Trello for collaboration with all participants in the pilot.	2018	J Fiddes
16	7	Engage communities of practitioners beyond those stations participating in the pilot project; use Twitter to promote the initiative, Identify teams to work on different components: dev, packaging, documentation, early adopters.	On going	J Fiddes; Ø Godøy
17	3.3	Data Portal Team to ensure that the Best Practices Guide reflects guidance on data quality	On going	Ø Godøy; GCW PM
18	3.3	GCW Observations WG to assess how to include the hydrological data, in addition to meteorological data, as ancillary information.	2018	W Schöner
19	3.3	Secretariat to communicate regularly with the Focal Points, and ensure that the terms of reference reflect their expected active engagement, including on the data and metadata exchange.	On going	GCW PM
20	4.	GCW to collaborate with ADC in the organization of the planned Arctic Data Forum in 2018, leading to a statement for the Arctic Science Ministerial	2018	Ø Godøy; GCW PM

21	5.	J Fiddes to involve CAS in the development of the application for data exchange pilot.	03.2018	J Fiddes
22	5.	GCW Secretariat to provide additional information regarding the linkages between CAS, CMA and the Polar Research Institute of China, and how these could help in improving the interoperability with the Data Portal.	GSG-5	GCW PM
23	6.	C Hrynkiw to assess the existing WIGOS and OSCAR metadata for cryosphere variable relative to the GCW observed variables, and report of differences, focusing first on those variables exchanged or would be exchanged on GTS (snow depth, SWE, etc.). Use Trello for collaboration (additional details in the report). Secretariat to work with C Hrynkiw to align the schedule of the Metadata work with the schedule of the Metadata TT meeting (Nov 27-29, 2017) and deliverables.	12.2017	C Hrynkiw; GCW PM
24	6.	GCW will continue to be engaged with related projects from other initiatives, like those in Europe, ADC-IARPC Vocabulary and Semantics WG, etc., regarding projects on terminology and semantics.	On going	Ø Godøy
25	7.	GCW to develop a strong documentation base to support all activities, as a form of capacity development, critical, in particular, to the small stations and data centres.	On going	GCW WG Leads
26	7..	GCW Data Portal to take into account the need for sharing, exchanging information as images, videos, etc.	On going	Ø Godøy
27	8.	CryoNet sites to provide to J Fiddes test cases and try with csv files, testing based on functionality in the system;	12.2017	S Marshall; T Kralidis; K Luojus
28	8.	Organize an informal follow-up at the Arctic Change Assembly 2017 (Quebec City). A presentation on GCW to be submitted for the conference, and to reference the work on Best Practice Guide development.	12.2017	A Walker; B Goodison
29	8.	B Goodison recommended that GCW leads the organization of a side event EC-70 on the exchange of info (snow depth, SWE).	06.2018	GCW PM

Annex 5: Acronyms

ADC	Arctic Data Committee
BUFR -	Binary Universal Form for the Representation of meteorological data
CAS	Chinese Academy of Science
CASRAI	Consortia Advancing Standards in Research Administration
CCIN -	Canadian Cryosphere Information Centre
CDM -	Common Data Models
CF -	Climate and Forecast (standard names and structures)
CKAN	Open Source data portal platform
CIMO	Commission for Instruments and Methods of Observations
DCPC	Data Collection and Production Centres
DOI	Difital Object Identifier
EC –	(WMO) Executive Council
EC PHORS –	WMO Executive Council Panel of Experts on Polar and High Mountains Observations, Research, and Services
ECCC -	Environment and Climate Change Canada
EnviDat	The Environmental Data Portal of SLF
GEO	Group on Earth Observations
GCW -	Global Cryosphere Watch
GSG –	GCW Steering Group
GCDM DIF	Global Change Master Directory – Directory Interchange Format
GPCs-LRF -	WMO Global Producing Centres
GRIB -	GRIdded Binary or General Regularly-distributed Information in Binary form
HMEI	Hydrometeorological Equipment Industry Association
IARPC	Interagency Research Policy Communication
INTAROS	INTAROS project, working towards an integrated arctic observing
INTERACT	International Network for Terrestrial Research and Monitoring in the Arctic
KPI	Key Performance Indicator
MetNo	Norwegian Meteorological Institute
NetCDF –	Network Common Data Form
NMHSs -	National Meteorological and Hydrological Services
OAI-PMH -	Open Archive Initiative Protocol for Metadata Harvesting
OPeNDAP -	Open-source Project for a Network Data Access Protocol
OGC- CSW	Open GeoSpatial Consortium – Catalog Services for the Web
OSCAR	Observing Systems Capability Analysis and Review Tool
PDC -	Polar Data Catalogue
PyWPS -	Python Web Processing Services
PR	Permanent Representative
PRCC	Polar Regional Climate Centre
RA –	(WMO) Regional Association
SAML	Secure Assertion Markup Language
SAON -	Sustained Arctic Observing Networks
SIOS -	Svalbard Integrated Arctic Earth Observing System
SLF	WSL Institute for Snow and Avalanche Research
SWE –	Snow Water Equivalent
WMO -	World Meteorological Organization
WIS	WMO Information System
WIGOS	WMO Integrated Global Observing System