



Technical documentation

*GCW Operational Manual for
contributing data centres*



Versions

Version	Date	Comment	Responsible
0.1	2015-11-25	First draft for internal discussion.	Øystein Godøy

Table of Contents

1	Introduction	4
1.1	Background	4
1.2	Scope	4
1.3	Intended audience	4
1.4	Applicable documents	4
2	Definitions	4
3	System overview	5
3.1	Introduction	5
3.2	Roles	5
3.3	Metadata	6
3.3.1	Background	6
3.3.2	Responsibilities	6
3.3.2.1	Contributing data centre	6
3.3.2.2	GCW Portal	6
3.4	Data	7
3.4.1	Background	7
3.4.2	Responsibilities	7
3.4.2.1	Contributing data centres	7
3.4.2.2	GCW Portal	7
4	Key performance indicators	7
4.1	Reporting	7
4.2	End point availability	7
4.3	End point response time	7
4.4	Number of datasets in the catalogue	8
5	Change management	8
5.1	Background	8
5.2	Procedures	8

1 Introduction

1.1 Background

The GCW Portal is the entry point to cryospheric datasets. It offers a web interface that contains information about datasets (metadata). These metadata are harvested on a regular basis from data centres actually managing the data on behalf of the owners/providers of the data.

The GCW Portal utilises interoperability interfaces to metadata and data in order to provide a unified view on the datasets that are relevant for GCW activities. The GCW Portal is also the interface for GCW metadata to WMO Information System (WIS) and WMO Integrated Global Observing System (WIGOS)¹. The GCW Portal will also facilitate real time access to data through Internet and WMO GTS² as requested by the user community. This requires a certain level of interoperability at the data level in addition to at the metadata level. On GTS WMO formats (BUFR and GRIB) is required and the GTS Portal can transform into these formats in the dissemination process provided contributing data centres are following the required standards for documentation and interfaces to data.

Guidelines to documentation and interfaces is provided in [2] .

1.2 Scope

This document provides an overview of the GCW Portal system and identifies aspects that has to be handled as well as key performance indicators that are required to operate the GCW Portal system in a sustainable manner.

1.3 Intended audience

System managers at the data centres contributing to the GCW Portal system. This includes data centres managing CryoNet data as well as other data centres managing cryospheric data. Requirements are stricter for data centres managing CryoNet stations data than for other contributing data centres.

1.4 Applicable documents

[1] [Global Cryosphere Watch \(GCW\) Implementation Plan, Version 1.5](#)

[2] GCW Portal Guidance for data centres contributing to GCW, Version 0.1

[3] <http://globalcryospherewatch.org/>

2 Definitions

CryoNet sites/stations - These are defined in [1] and [3] ³.

End point - An end point in GCW Portal context is a service URL being utilised within the

1 Details on how to avoid duplicate information in WIS and WIGOS needs to be defined.

2 For datasets not routed through GTS by other agencies.

3 See http://globalcryospherewatch.org/cryonet/site_types.html

GCW Portal system. Some end points have multiple purposes. E.g. THREDDS may offer HTTP, OGC WMS and OPeNDAP end points for datasets. In this context monitoring of a single end point URL is sufficient.

3 System overview

3.1 Introduction

GCW Portal is a metadata driven system where metadata describes the datasets and the interfaces available for the specific dataset. Metadata are collected in a central metadata catalogue while the data is served directly from the individual data repositories contributing to the GCW Portal system. An overview of the GCW Portal metadata flow is provided in Figure 1. This is under continuous development and should only be considered as an illustration.

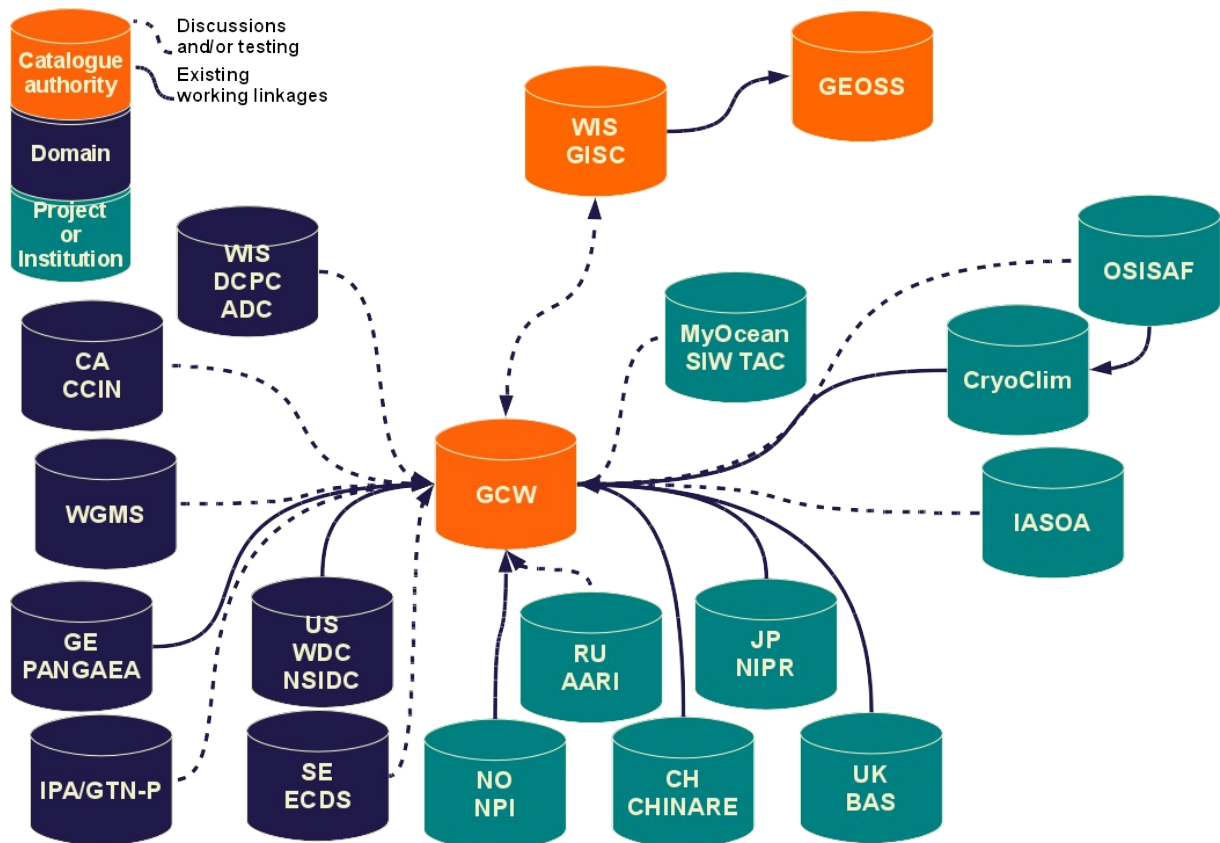


Figure 1: Schematic overview of contributing data centres in the GCW Portal. Some linkages are working, some under testing and some planned.

3.2 Roles

In relation to the GCW Portal a number of roles are usually defined. These are required to operate a data centre, to integrate across data centres and to fill the data centres with actual content.

Data manager – The data manager is responsible for the content of the data centre. The data manager guides data providers in the process of documenting and formatting data as

well on submission. The data manager is responsible for the life cycle management of data (both real time dissemination and long term preservation). The data manager is responsible for interaction with other data centres and with the GCW Portal⁴.

System manager – The system manager is responsible for the technical framework used to manage and communicate data. The system manager is responsible for the toolbox used by the data manager, data provider and data consumer.

Data provider – The data provider is the Principal Investigator responsible for the dataset being made available from a station contributing to GCW. A data provider may be responsible for one or more datasets and relates to one or more data managers depending on the routing (real time and long term preservation) of datasets.

Data consumer – The data consumer is a user of datasets through the GCW Portal (whether archived or real time). The data consumer depends on the work performed by the data provider, data manager and system manager in order to find (through discovery metadata), access (through interfaces to data) and use (through use metadata fully describing the content of a dataset) relevant datasets.

3.3 Metadata

3.3.1 Background

Metadata are generated by the data centres hosting the data sets. Metadata are harvested and ingested in the central catalogue for usage by the GCW Portal user community. For information of various types of metadata and exchange mechanisms the reader is referred to [2].

3.3.2 Responsibilities

3.3.2.1 Contributing data centre

1. The contributing data centres are responsible for setting up metadata endpoints exposing metadata according to the GCW guidelines [2].
2. The contributing data centres are responsible for informing the GCW Portal on interruptions in the availability of the endpoints.
3. The contributing data centres are responsible for applying the RFC (Change management) procedure when changes are required to the local system.
4. The contributing data centres are responsible for the quality of the metadata they are exposing.

3.3.2.2 GCW Portal

1. The GCW Portal will twice daily access metadata and points and harvest the metadata that has changed since the last harvest. At regular intervals the content of the GCW Portal will be compared with a full harvest.

⁴ Data managers of contributing data centres will be part of the GCW Portal Team.

2. The GCW Portal is responsible for mapping the acknowledged metadata standards to the search model used by the GCW Portal.
3. The GCW Portal is responsible for applying the RFC procedure (Change management) when changes are required in the central system.

3.4 Data

3.4.1 Background

The data are normally not served by the GCW Portal but directly from the responsible data centre. However, the GCW Portal will, where possible through interfaces announced by the data centre facilitate combination of datasets or transformations if possible⁵.

Furthermore, the GCW Portal will interface with WMO GTS for relevant datasets.

3.4.2 Responsibilities

3.4.2.1 Contributing data centres

1. The contributing data centres⁶ are responsible for the quality of the data exposed.
2. The contributing data centres are responsible for the availability of the end points announced for the datasets.

3.4.2.2 GCW Portal

- 1.

4 Key performance indicators

4.1 Reporting

The information collected will be reported to GCW Steering Group.

4.2 End point availability

As GCW Portal is a distributed system, it is fully dependent on the availability of service end points distributed across GCW Portal contributing data centres. The GCW Portal will monitor the availability of registered end points for metadata and data.

Although GCW Portal is monitoring the availability of end points, it is the responsibility of each contributing data centre to ensure the availability of their end points. If GCW Portal actions on endpoints are found to break functionality, these must be documented and reported.

4.3 End point response time

The user experience depends on the response time of the end points used. The GCW Portal relies on metadata harvest so the response of end points will not affect the user experience

5 This will normally require data to be available through OPeNDAP services.

6 In cooperation with the Principal Investigators.

when search for data, but when the user tries to access data this relies on the response time of the end points at the contributing data centres.

Currently no specific time limit is set and a proposal for such will be provided after monitoring of the system in real use for a period.

4.4 Number of datasets in the catalogue

The number of datasets documented harvested and deleted by the GCW Portal will be monitored and reported. This information can be segmented on CryoNet sites/stations provided each CryoNet site/station only has one end point for metadata. This is however not the situation for all and some mechanism for handling this is required to be developed.

5 Change management

5.1 Background

As GCW Portal is a distributed system it relies on the interaction of a number of subsystems physically distributed between different organisations. If one organisation makes changes to their subsystem without sufficient prior notice, the GCW Portal functionality breaks.

The the GCW Portal must implement a mechanism for change management, but the approach is to do this in a pragmatic manner⁷.

5.2 Procedures

- 1 All data centres contributing to GCW Portal has to create a Request for Change (RFC) when planning modifications of subsystems. These RFCs are sent by email to gcw@met.no⁸.
- 2 The Norwegian Meteorological Institute, being responsible for the catalogue interfaces, is responsible for informing all affected parties⁹, collecting responses from these and creating a RFC progress plan.
- 3 The normal life cycle for an RFC involves the following steps, where the first 3 should take no more than 1 month.
 - 3.1 Request
 - 3.2 Announcement
 - 3.3 Progress planning
 - 3.4 Implementation
 - 3.5 Evaluation
 - 3.6 Closure

⁷ At least until more formal and bureaucratic procedures are required.

⁸ Implementation of request tracking using RT, Redmine, Phabricator, Trello or similar are under consideration.

⁹ Many modifications will only affect the relationship between the catalogue and the contributing data centre.