Data Catalogue
Status and demonstration

Øystein Godøy
Purpose

● To provide an overview of datasets relevant to GCW
● To provide access to datasets wherever possible
   • Real time data streams
   • Archive access
● To connect GCW with WMO Information System
● Distributed Data Management
   • Metadata driven
   • Currently not hosting data
Issues that need special care

• Bridging scientific and operational communities
  • e.g. WMO WIS ↔ ICSU WDS
  • including data centres not belonging to a framework

• Machine interfaces
  • Interoperability standards
    – Compliance
    – Utilisation
Metadata

• Discovery metadata
  • Who measured what, when and where?
  • Where to get data?
  • How can data be used?
  • Link to auxiliary information like product manuals, …

• Use metadata
  • Required to use the data
    – Detailed information on parameters provided
    – Units, missing value coding, map projections, …
  • Accessed when dataset is found
Changes since last meeting

• The demonstration catalogue and portal has been shut down
  • The old address points to new server
• A new catalogue and portal has been implemented
  • http://gcw.met.no/
  • Initialisation of new database autumn 2013
  • Regular harvest of metadata
  • New hardware
  • Updated software
Status (1)

- Manually initiated metadata harvest ("push button")
  - National Snow and Ice Data Centre
  - British Antarctic Survey
  - CryoClim

- Have tested against
  - CCIN
  - WGMS
  - PANGAEA
  - NPI
  - (EUMETSAT)
Status (2)

- Automatic filtering of harvested records
  - No dedicated “cryosphere” sets
  - Filters on keywords related to cryosphere
- Manually supervised ingestion of harvested metadata
  - Reasons to be explained
  - Formats used so far GCMD DIF, ISO19115, FGDC, ad hoc formats, …
- Metadata are exposed through OAI-PMH
  - GCMD DIF and ISO19115
Issues that require attention (1)

- Filtering of harvested metadata
- Usage of controlled vocabularies
  - Machine readable implementations are required
- Brokering and cleaning/validation of metadata
  - For WIS compliance
  - WMO controlled vocabularies (and consistency of these)
- Duplication of metadata
  - Which metadata to expose through machine interfaces
- Metadata stewardship
Issues that require attention (2)

- Ambition level for higher order services
  - Visualisation, transformation etc
- How to develop and propagate controlled vocabularies towards other parts of WMO and externally
- Similar appearance of website and data catalogue human interfaces?
- How to identify contributing data centres?
- Approval of some datasets?
http://gcw.met.no/