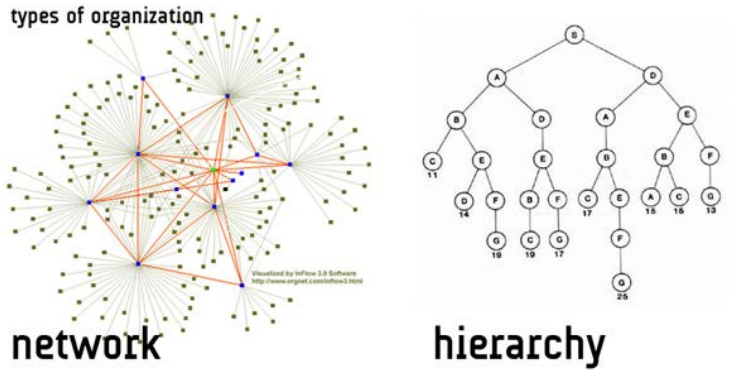
Arctic Data Centre

Bridging scientific and operational data streams

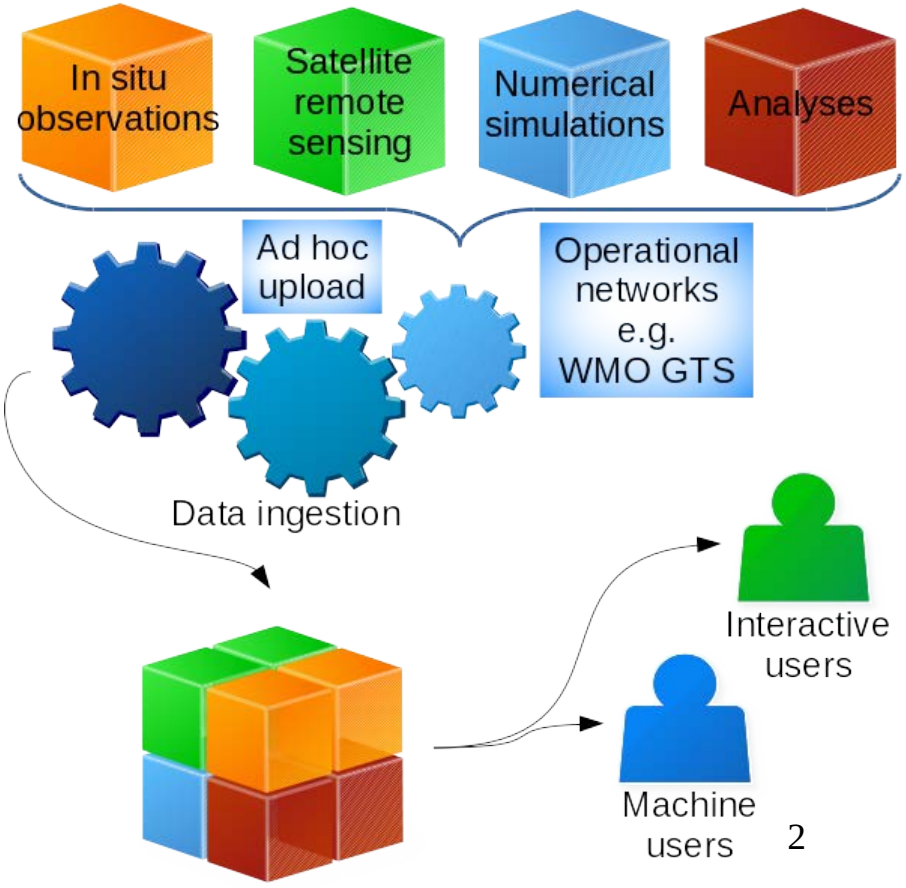
Øystein Godøy and Bard Saadatnejad



From http://n-e-r-v-o-u-s.com/blog/?p=26

Background

- A legacy of IPY
  - International Operational Data Coordination through IPYCOORD
  - National Data Coordination through DOKIPY
- WMO Information System
  - Data Collection and Production Centre



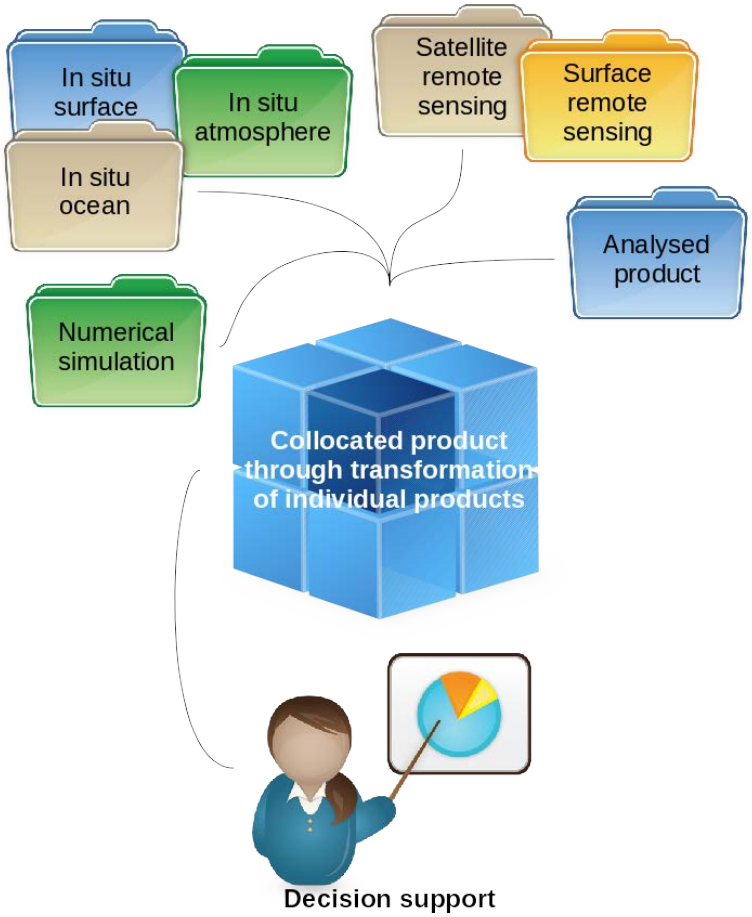
Supporting projects

- WMO
  - Global Cryosphere Watch
  - Year of Polar Prediction
- Research Council of Norway
  - Norwegian Satellite Earth Observation Database for Marine and Polar Research (NORMAP)
  - Norwegian Marine Data Centre (NMDC)
  - Norwegian Scientific Data Network (NorDataNet)
  - Svalbard Integrated Arctic Earth Observing System (SIOS KC)
  - Pilot project for geoscientific data management (GeoAccessNO)
- EU FP6/FP7/H2020
  - DAMOCLES
  - ACCESS
  - APPLICATE
  - INTERACT
- EUMETSAT
  - Ocean and Sea Ice SAF
- Copernicus
  - Copernicus Marine Environmental Monitoring Service
- ESA/Norwegian Space Centre
  - CryoClim
  - National Ground Segment for Copernicus data

Relation between ADC and projects

- ADC is the long term data management commitment focusing on both scientific and operational data streams of the Norwegian Meteorological Institute
- All data hosted in project specific portals will be included in ADC
- Functionality targeting the scientific user community is normally developed through dedicated projects

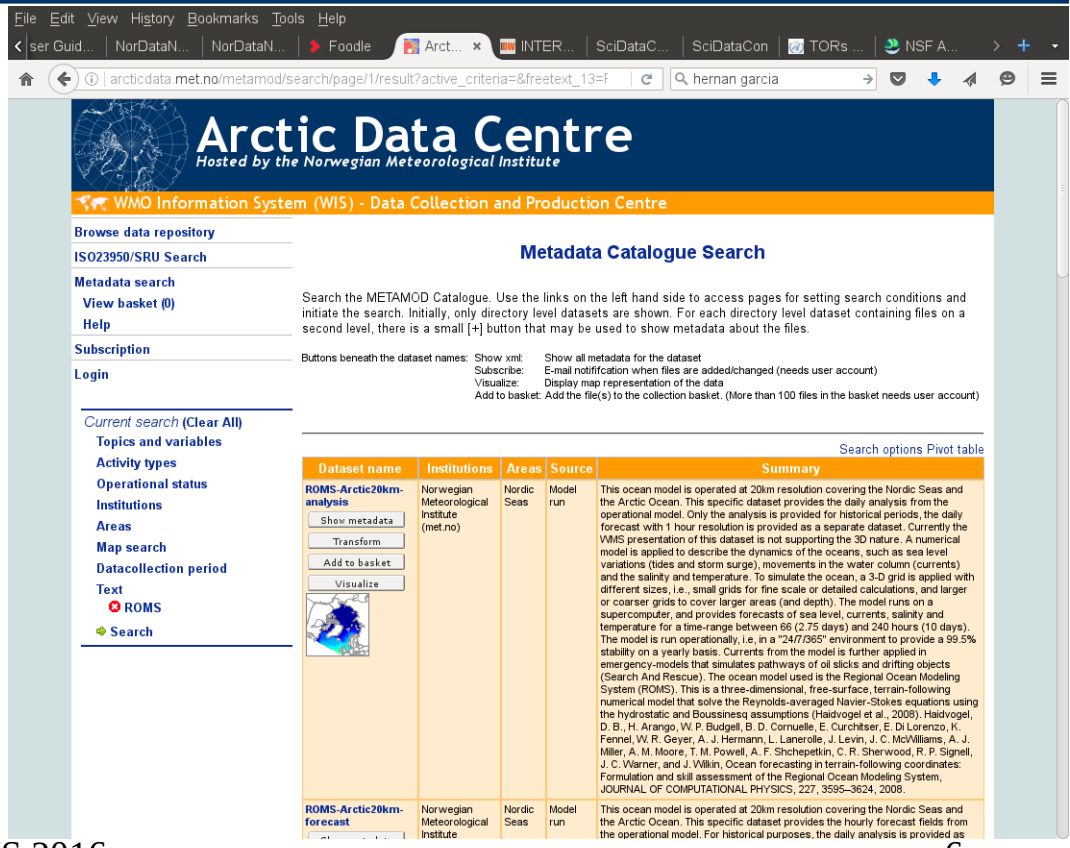
Approach (1)



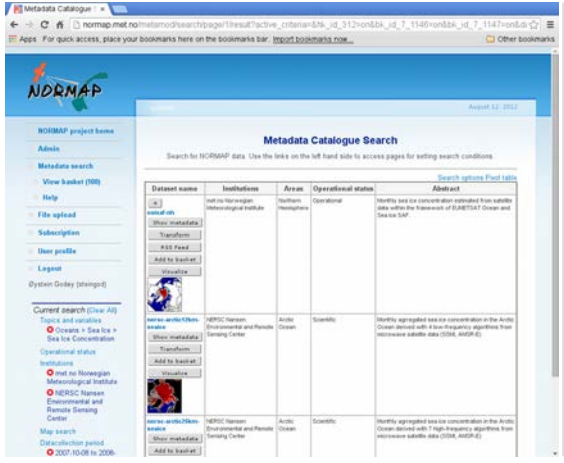
- Dataset oriented
  - Metadata driven
- Open data space
  - Higher order services offered when the data space can be constrained
- Net centric
  - Linkages with other data centres is vital
  - Implies brokering of metadata and data
- Interdisciplinary
  - Dataset agnostic in the open data space

Approach (2)

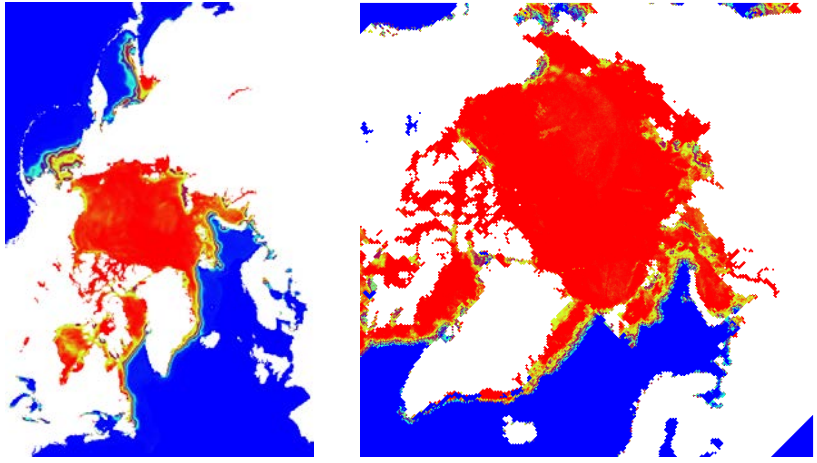
- Technical development through METNO Scientific Information System (METSIS)
- Not a necessarily a distribution system
- Rather a toolbox
  - To set up distribution services for projects and the general public
  - Easily configurable frontend to credit various projects
  - Sharing backend elements for cost effective implementation, operation and maintenance
- Utilisation of open source software where possible
- Releasing software developed as open source code (GPL/LGPL)

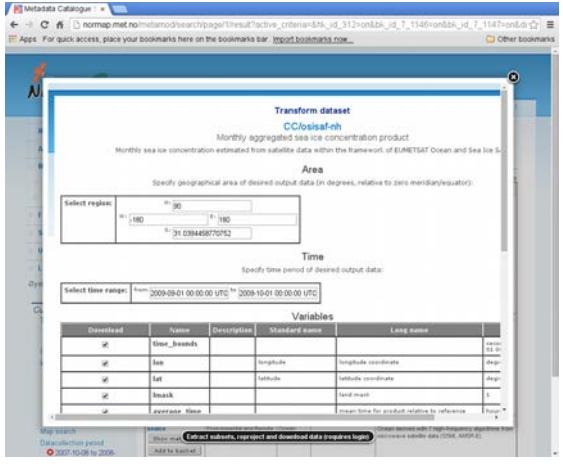


Transformations allow users to do comparisons of products and to extract tailored products for their specific need

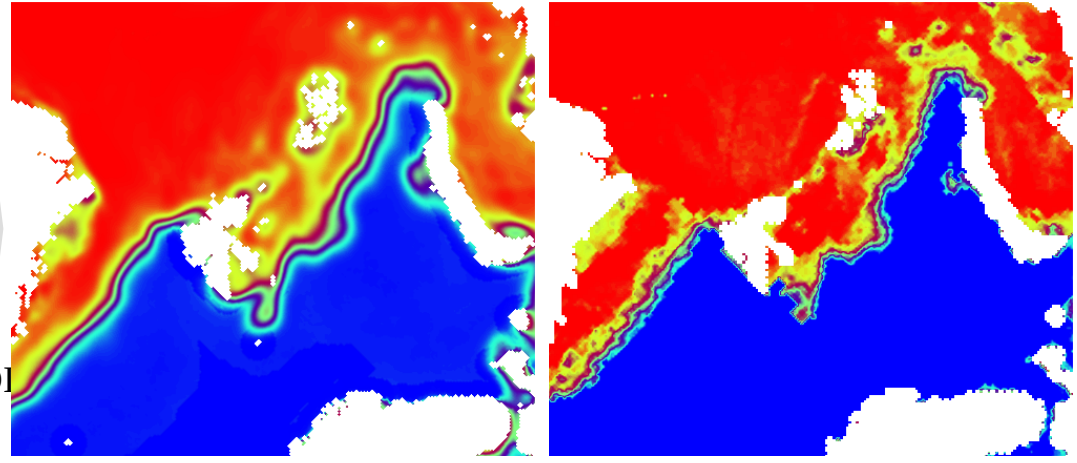


Search results





Transformation request



User interaction

- Primarily through supporting projects
  - Presentations at general assemblies through science oriented EU projects
  - User workshops through national e-infrastructure projects
- User oriented articles in the national e-infrastructure magazine META
- Questionnaires
- Polls and issue tracking for feature requests
- Much the same people responding regardless of method
- Increasing awareness in the scientific community as funding agencies are starting to require data deposition

Lessons learned

- Flexible and adaptable user interfaces are vital
  - to establish a good dialogue with users
- Basket/Cart is required
  - to move from open to closed data space enabling higher order services
- Data streams instead of file transfers reduce complexity and cost
  - Reducing housekeeping requirements
  - Especially important for data brokering
- Interdisciplinary translation of documentation structure is “easy”
  - but translation of content is difficult
- Machine readable controlled vocabularies are needed to bridge communities
  - Blocking issues are more cultural than technical
- The requirement for use metadata is underestimated
  - Standardised data documentation as in NetCDF/CF and conveyed through OPeNDAP is highly valuable for automated interaction between data centres
- Configuration metadata are required in distributed data management
  - Should preferably follow index metadata

Future plans

- Currently Developing a new human frontend
  - Standard CMS using Drupal 7
    - functionality as web services
  - More flexible to serve user requirements and editorial content
- Currently also modularising software
  - To ensure flexibility and sustainability when responding to user requirements
- Using OGC WPS as communication layer
- Work flow management
  - Linked to HPC services
- Extended semantic translations
- Efficient utilisation of DOI in KPI reporting
- Integration with the national high performance data archive
  - To support interdisciplinary use and simplify integration with national HPC services