Data Management in Science and the Legacy of the International Polar Year

- Ellsworth LeDrew
- Mark Parsons
- Taco de Bruin

- Peter Yoon
- Christine Barnard
- Scott Tomlinson
- Warwick Vincent

  - http://www.earthzine.org/2008/03/27/securing-the-legacy-of-ipy/

The Role of Data management In IPY

- IPY 1 1882: much of the data has been lost
- IPY 2 1932-33, almost nothing remains
  http://www.arctic.noaa.gov/aro/ipy-1/index.htm
- "Building an integrated data set from the broad range of IPY research activities represents one of IPY's most daunting challenges. An enduring data set, accessible to scientists and the public during IPY and for many decades into the future, will represent one of IPY's strongest legacies "
  (The Scope of Science for the International Polar Year, 2007-2008, http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Inform

Metadata

- In accordance with the ISO standard Reference Model for an Open Archival Information System (OAIS) (CCSDS 2002), complete metadata may be defined **as all the information necessary for data to be independently understood by users and to ensure proper stewardship of the data**. Regardless of any data access restrictions or delays in delivery of the data itself, all IPY projects must promptly provide basic descriptive metadata of collected data in an internationally recognized, standard format to an appropriate catalog or registry.
- Metadata are essential to the **discovery**, access, and effective use of data.
- All IPY data must be accompanied by a full set of metadata that completely document and describe the data.

MetaData

- ISO: International Organization for Standardization (http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020)
- FGDC: Federal Geographic Data Committee (http://www.fgdc.gov/policyandplanning/fgdc-policies)
- GCMD: Global Change Master Directory (http://gcmd.gsfc.nasa.gov/KeywordSearch/Home.do?Portal=GCMD_Services&MetadataType=1)

Data Access

Implicit in the concept of metadata is that colleagues should have access to your metadata

- "IPY data, including operational data delivered in real time, are made available fully, freely, openly and on the shortest feasible timescale. Exceptions will only apply to protect confidentiality of information about human subjects, respect needs and rights of holders of local and traditional knowledge and ensure that data release does not lead to harm of endangered or protected resources."
  (http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/1155_DD_FILE_IPY_Science_Plan.pdf)
- "shortest feasible timescale": allow time for basic Validation and QC, of order of months, not years.

Data Acknowledgment

placing a higher value on the publication of data

- To recognize the valuable role of data providers (and scientists who collect or prepare data) and to facilitate repeatability of IPY experiments in keeping with the scientific method, users of IPY data must formally acknowledge data authors (contributors) and sources.
- Where possible, this acknowledgment should take the form of a formal citation, such as when citing a book or journal article.
- Journals should require the formal citation of data used in articles they publish. Where formal citation is not possible, such as with some medical and social science data, ethical policies for data collection and data use are encouraged, building upon existing models such as Article 8(j) of the 1992 Convention on Biological Diversity.

Data and Information Rescue

- Scientists should inventory major collections of extant data and information and should set priorities for the rescue and permanent preservation of the data and information that are most valuable and at greatest risk.

Professional Data and Information Management

- Financial support for data and information management should become a routine component in all research budgets and the evaluation criteria for assessing research funding proposals should include evaluation of data management.
- All scientists should receive training in data management as part of their graduate and postgraduate education.
- Scientists should be recognized and given credit for the scientific contribution of the data sets that they produce as well as for the analysis of those data.

OnGoing Issues

- Interoperability - GEOSS
- Equitable Access to Data and Information (Open Access for Data and Publications)
- The Digital Divide
- Intellectual Property Rights
- Government held data vs Private Sector Data
- Role of World Data Centres and Monitoring Centres (GOOS, GTOS, GCOS)

A collaboration between CCIN, ArcticNet, IPY, and NCP (others in development): http://www.polardata.ca.

- Support the Capture of metadata through a simple interface using international standards
- Seamless Discovery Operation with other Portals
- AHelp Desk:
  - Guide Metadata Input
  - Inform, educate & support throughout
  - Support Privacy assessments & other legal obligations
- Archive data and derived products for specific projects
- Develop Best Practices Wiki
- Develop outreach and educational products
- Design interface using approaches and tools familiar to local peoples.

HTTP://WWW.EARTHOBSERVATIONS.ORG

GEOSS: The Global Earth Observation System of Systems

Click on Societal Benefit Area icons to go to the relevant page

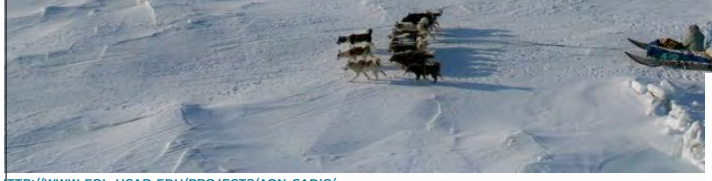
ACCESS. CONNECTING. USERS.

The Global Earth Observation System of Systems will provide decision-support tools to a wide variety of users. As with the

The Arctic Observing Network (AON) Cooperative Arctic Data and Information Service (CADIS)

Florence Fetterer, James Moore, Don Middleton
Mohan Ramamurthy, Michael Burek, Don Stott, Jeff Weber

Second AON PI Meeting
Arctic Observing Integration Workshops
Palisades, NY
17-20 March 2008



Cooperative Arctic Data and Information Service of the ARCTIC OBSERVING NETWORK
CADIS

http://www.eol.ucar.edu/projects/aon-cadis/

Primary CADIS Goals

- CADIS is the AON Data Portal that will evolve
- Support NSF IPY AON data streams
- Strive to develop end-to-end data services
- Prototype data discovery, ingest and distribution capabilities
- Emphasis on developing a unified metadata standard that is compatible with ISO, IPY and other international conventions
- Build a consensus data set format that will allow one to display, subset and analyze AON network data

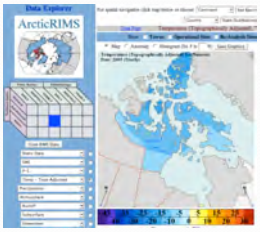
Cooperative Arctic Data and Information Service of the ARCTIC OBSERVING NETWORK
CADIS

http://www.eol.ucar.edu/projects/aon-cadis/

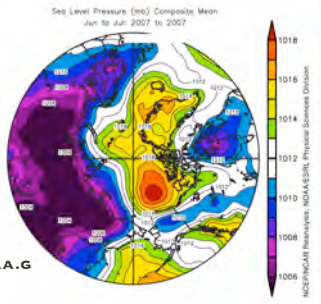
Striving to Meet PI and Community Expectations

- Enter metadata quickly, efficiently, and without confusion
- Find my own data and display it
- Display it over or with other data sets of interest (that may or may not be in the CADIS system)
- Mathematically manipulate data
- Export in a choice of formats
- Easy to use, graphical interfaces

Some examples to emulate



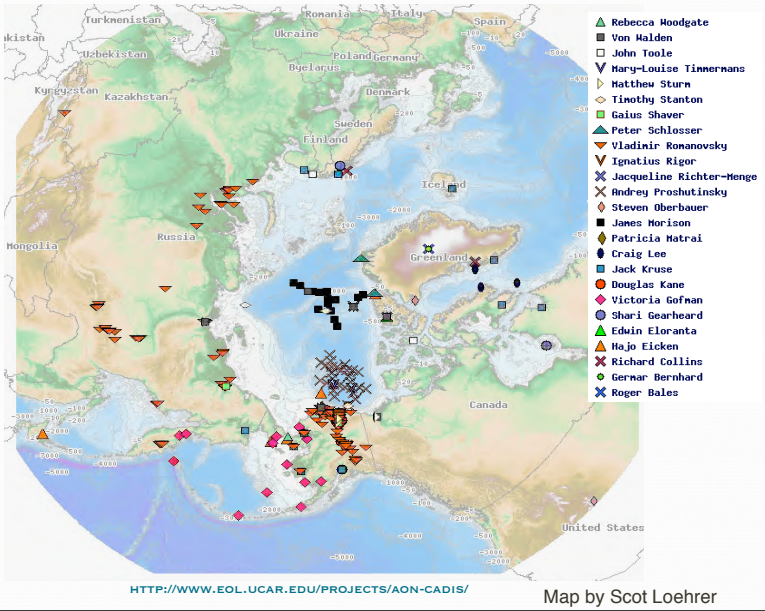
ArcticRIMS - http://rims.unh.edu



NOAA ESRL PSD - http://www.cdc.noaa.gov/PublicData/

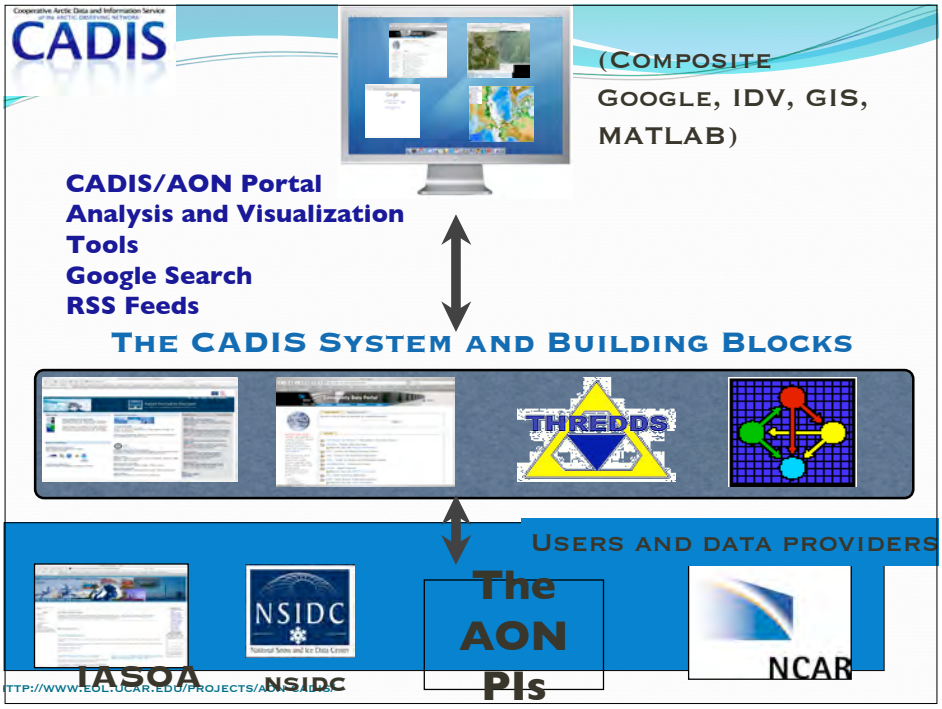
http://www.eol.ucar.edu/projects/aon-c

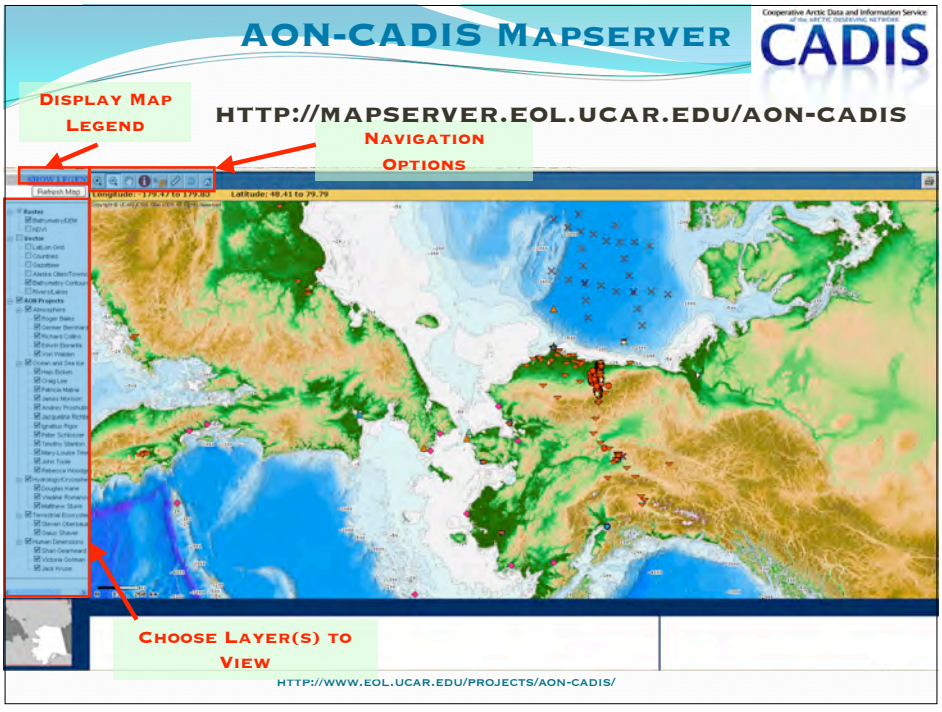
AON Site Pan-Arctic Distribution



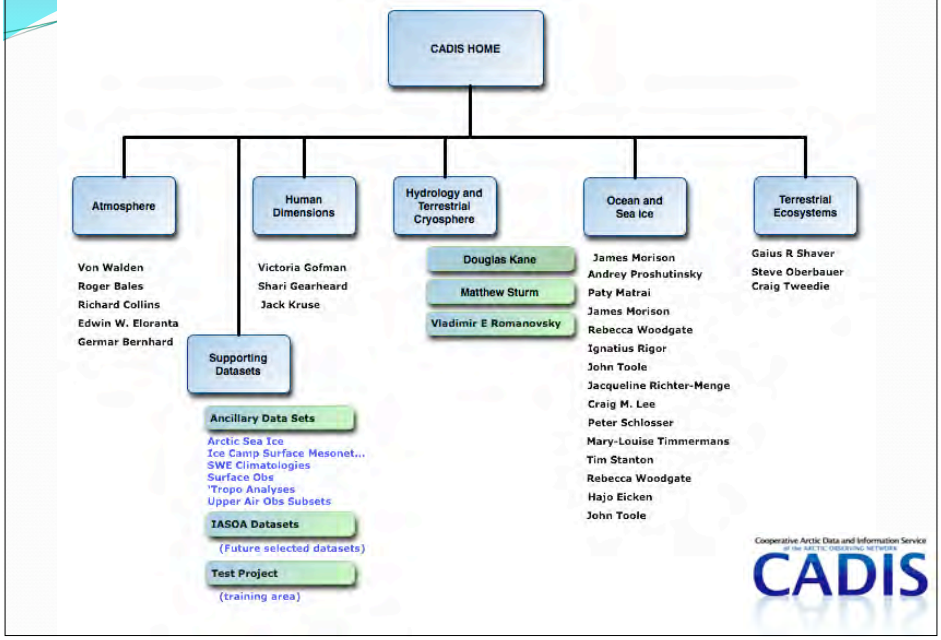
http://www.eol.ucar.edu/projects/aon-cadis/

Map by Scot Loehrer

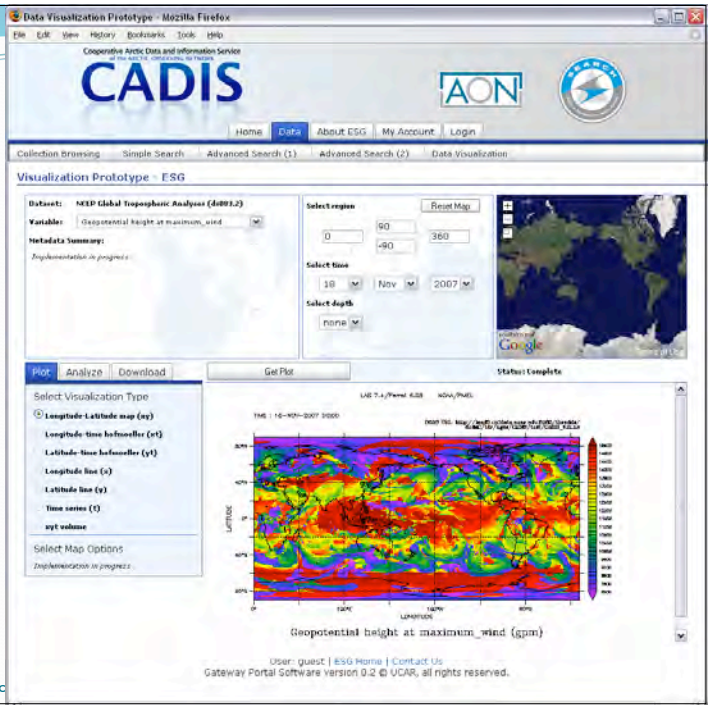

CADIS
(Composite
Google, IDV, GIS,
MATLAB)
CADIS/AON Portal
Analysis and Visualization
Tools
Google Search
RSS Feeds
The CADIS System and Building Blocks
THREDDS
Users and Data Providers
NSIDC
The
AON
PIs
NCAR
http://www.eol.ucar.edu/projects/aon-cadis


AON-CADIS Mapserver
CADIS
Display Map
Legend
http://mapserver.eol.ucar.edu/aon-cadis
Navigation
Options
Choose Layer(s) to
View
http://www.eol.ucar.edu/projects/aon-cadis/

Metadata and Data Portal



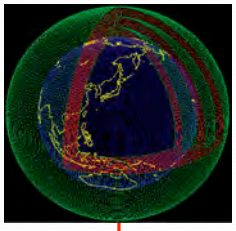
Analysis and Visualization –Live Access Server

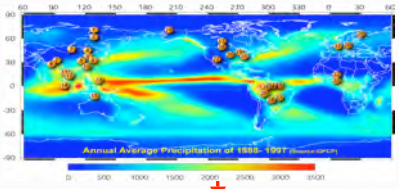


HTTP://WWW.EOL.UCAR.EDU/PROJEC

Coordinated Energy and water-cycle Observations Project

A Well Organized Data Archive System





In-Situ Data Archiving Center at NCAR (National Center for Atmospheric Research) of USA

Model Output Data Archiving Center at the World Data Center for Climate, Max-Planck Institute for Meteorology of Germany

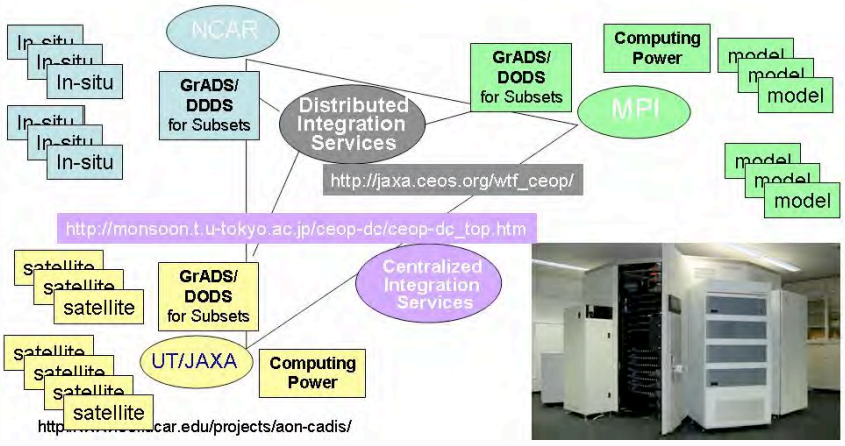
Data Integrating/Archiving Center at University of Tokyo and JAXA of Japan

http://www.eol

Coordinated Energy and water-cycle Observations Project

Data Management

Distributed- and Centralized- Data Integration Functions



Suggested activities from IGOS report on Cryosphere:

- Work with the CEOP (Coordinated Energy and water cycle Observations Project) project to further develop tools for integrating diverse and geographically distributed remote sensing and in situ data.
- Work with IPY DIS to find suitable permanent archives for IPY data.
- Collaborate with data centres and research projects in the application of standards and techniques for data and information management.
- Identify gaps in climate and cryospheric data requirements and seeking opportunities to address these gaps (it is expected the DIS will identify several of these gaps).
- Encourage and initiate data collection and recovery projects.
- Encourage the reprocessing of data sets for climate studies.
- Identify data archives that are not open for easy access, and promote the use of standards to improve access.
- Advise individual, collaborating projects on data and information management issues.

Issues from Discussions - One

- Do not reinvent the wheel, use existing systems if possible
- Network of networks, including linkages to extra-Arctic nodes and supporting data sources
- Incorporate local knowledge and the system should be of use for local people - a two way process
- Incorporation of Reanalysis data and experiments
- Appropriate access of information: real time monitoring of caribou collar signals by hunters is NOT appropriate
- Keep politics out of process
- Data Push vs Data Pull process

Issues from Discussions Two

- **Data** versus **Knowledge** of value to users: information commons
- Provide for transition from science to operations
- Provide an interface for public and policy makers
- Issues of intellectual property and ownership vs collaboration
- Real time products, such as for weather or environment channel

Issues from Discussions Three

- Information Management (reports, publications, etc.)
- Data Stewardship
- Long term data/metadata access
- Must show benefits to the Data Providers
- Importance of Metrics

Summary of Cyberinformation and Data Management - Major Issues

- Do not reinvent the wheel, use existing systems if possible
- Network of networks, including linkages to extra-Arctic nodes and supporting data sources
- **Data** versus **Knowledge** of value to users - Education and Outreach component – Web 2.0
- Provide for transition from science to operations
- Recognition of data – peer review and citation
- Issues of intellectual property and ownership vs collaboration, includes traditional knowledge.
- Incorporation of Reanalysis data and experiments, and model products

Next Steps

- Begin discussion and coordination among all potential distributed arctic data systems/portals (establish Cyberinformation and Data Management Working Group)
- Verify that Discovery Metadata is interoperable
- Update Inventory of all networks/data sources
- Work closely with WIS and GEOSS developments
- Need to develop SAON Data Policy