Thirty Years of Data Management for Earth Observations at the National Snow and Ice Data Center (NSIDC)

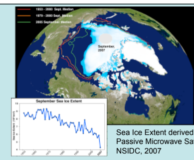
Roger G. Barry, Ronald L. S. Weaver, Richard L. Armstrong, Florence Fetterer
http://nsidc.org

Data Preservation Without Access is Pointless
Data Access Without Preservation is Impossible
Overheard in the NSIDC halls, 2004

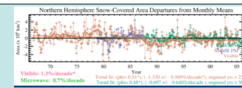
Some simple truths about observing systems emerge from our history of storing and serving data

New instruments have to be backward compatible with old

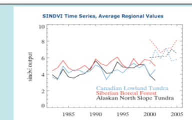
- A few long, continuous records are more valuable than a few short ones



Sea Ice Extent derived from Passive Microwave Satellite Data NSIDC, 2007



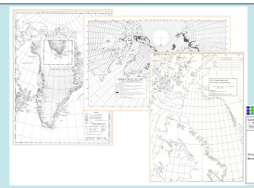
Northern Hemisphere snow covered area (SCA) departures from monthly means, 1978-2006, from NOAA snow charts (orange) and microwave satellite (purple/green) data sets. The NOAA time series for this period exhibits a significant decreasing trend of -1.3% per decade (solid orange line); the microwave snow cover time series exhibits a decreasing trend of -0.7% per decade that is not significant at a 90% level (dashed green line). (Brodzik et al. 2006)

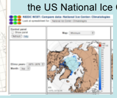


Seasonally integrated NDVI (greenness) from AVHRR (solid lines) and from MODIS (dashed). The MODIS product is more accurate, but if used alone the time series is too short to show trends. (http://nsidc.org/noaa/search/indicators/green_sindvi.html)

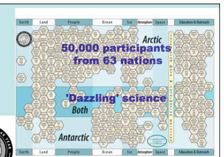
International collaborations are essential for building networks and for establishing data sharing protocols

- **Sea Ice Mapping:** Global Digital Sea Ice Data Base (GDSIDB)
- **International Polar Year:** IPY Data and Information Service (IPYDIS)
- **The Cryosphere:** Global Cryosphere Watch (WMO)
- **Glacier Mapping:** Global Land Ice Measurements from Space (GLIMS)





The GDSIDB helping coordinate digitization and archive format standards for operational ice charts. Through this work older paper charts like those from Denmark, the U.S., and Canada (left) can be combined with more recent charts created in a GIS environment, and further processed to make gridded data sets and climatologies useful for research, like the US National Ice Center chart climatology (below).



The International Polar Year 2007-2008 provides a framework for integrative science

Strategic IT decisions impact the cost and effectiveness of systems for managing and distributing observing system data

The challenge to IT managers is to sustain levels of service while driving costs downward through balanced use of technology and staff resources

From a Vertically Integrated "stovepiped" system of Today

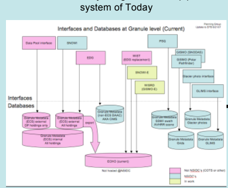
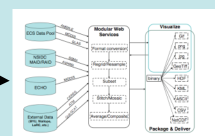
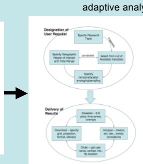


Diagram of the Current NSIDC Data System

To a modular "service bus" model



Allowing interactive data set exploration and adaptive analysis by the Scientist

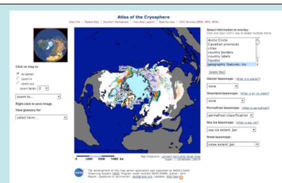


Conceptual diagrams from NSIDC's recently funded Cryospheric Change Analysis Web Services Proposal (NASA)

Prototype Mockup
Users can select from a number of base maps; toggle which data layers are displayed; browse time-series of a parameter over a region; and produce time-series plots and scatter diagrams. Users can download data for off-line analysis specifying the grid, format, projection, and delivery mechanism from the suite of available options.

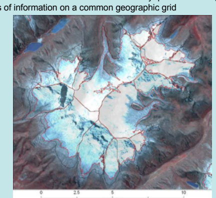
While some observations are useful on their own merit, many are only informative when combined with other like or contrasting observations

- Emerging GIS technology allows for integration of spatial data
- Virtual globe technology allows rapid visualization of diverse data sets
- Combinations of data and models lead to new conceptual understanding

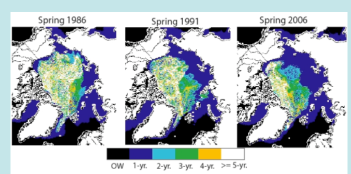


Screen shot from the NSIDC Atlas of the Cryopshere showing several layers of information on a common geographic grid

Placing data in 3D visualizers such as Google Earth provides a method to rapidly present data to the science and lay user communities. In this image (left) NSIDC sea ice extent derived from passive microwave data and permafrost extent derived from various sources are 'draped' over the Google Earth globe.



An example of glacier change analysis from the GLIMS project at NSIDC. Results show a decrease in glacier area in the Cordillera Blanca (Peru) of 15% over the period 1970 (black line, from map data) to 2003 (red line, from satellite data. (Racoviteanu, 2007)



No direct methods for the measurement of sea ice thickness over the entire Arctic Basin yet exists. This estimation is derived from Cryosat satellite altimetry data coupled with lagrangian tracking of ice motion from AVHRR data. (image courtesy Maslanik et. al., in press)

The National Snow and Ice Data Center is located at the University of Colorado, Boulder in the Cooperative Institute for Research in the Environmental Sciences.

NSIDC provides data services for most areas of the Cryosphere including sea ice,global and regional snow cover, ice sheets, glaciers, permafrost and freshwater ice. Our holdings are primarily derived from satellite data.

The authors acknowledge the continued strong support of the NSF, NOAA, and NASA

Our Data Management "Mantra"

Scientists must actively participate in data management

Successful Earth science data centers incorporate scientists in the data management process.

Scientists on staff help guide in-house product and tool development.

Scientists who use the data help improve data quality and weed out the inevitable 'bad data' points.

Data sets are scholarly products in their own right. They should be cited when used. NSIDC provides handy citations for doing this.

Scientists need rapid access to Earth science data and data centers must:

Exploit internet technologies (hypertext, distributed data catalogs, grid computing, virtual observatories, data portals, data mining etc.) to help scientists find data.

Produce climate data record quality time series in near real time.

Be ready to reprocess data based on newer algorithms.

Science and technology keep changing

Data and documentation must be durable enough to withstand constant technological change. (Paper has proven to be a more reliable archive medium than many digital formats).

Data management science is currently debating, and slowly accepting, reference standards for data management systems and procedures.

Data systems and especially the technology they use continue to evolve at a breakneck pace -- change is the constant.

New technologies bombard data system developers and operators --Hard to choose up front the long-lived technologies suitable for archiving data.

Data management includes planning for permanence

Research projects that consider where data will reside over the long term help assure that the important data are transferred to a Long Term Archive (LTA) facility accurately, completely and cost-effectively.

Data centers are not always equipped to be true LTA facilities (for one reason they lack guaranteed funding) but attention to metadata goes a long way toward assuring that data can be used in the distant future.

Requirements for LTA or preservation metadata go beyond descriptive information requirements.

Data producers and data users view data differently

Data producers (especially those involved in remote sensing) are interested in regularized production of data sets. Generally these data are timeseries, designed for a specific observational purpose.

Data users often want to subset, superset, or reprocess the data producer's output.

Data centers help by providing easy to use tools for manipulating data. BUT tools that work with all data sets don't exist, because data come in many different formats.

Data managers and scientists must jointly find acceptable formats and levels of metadata detail for each data set.

Data management takes time and money

The NRC estimates that 10% of a project's budget should be reserved for data management.