National Snow and Ice Data Center
A brief overview and data management projects

Walt Meier, Mark Parsons, and Mark Serreze
The National Snow and Ice Data Center...

- Manages and distributes scientific data
- Performs scientific research
- Creates tools for data access
- Supports data users
- Supports data users
- Educates the public about the cryosphere
**Products**

More than 600 data and information products, most freely available online

- Satellite
- In situ (station data and the like)
- Model output
- Most digital, some analog

**Users**

- Research/Higher Ed: 48%
- Government: 11%
- K-12: 3%
- Other: 27%
- Commercial: 9%
- Press: 2%
Funding agencies

- NASA: 71%
- NSF: 20%
- NOAA: 6%
- Other: 3%

Objectives

- Data management: 80%
- Research: 20%
NSIDC Distributed Active Archive Center

Data from NASA’s past and current Earth Observing System (EOS) satellites and other satellite and field measurement programs.

<table>
<thead>
<tr>
<th>Passive microwave</th>
<th>VIS/IR Moderate Resolution</th>
<th>Satellite &amp; Airborne Altimetry</th>
</tr>
</thead>
<tbody>
<tr>
<td>• AMSR-E (Aqua)</td>
<td>• MODIS (Terra/Aqua) snow and ice products</td>
<td>• ICESat I / GLAS altimetry and atmospheric lidar data</td>
</tr>
<tr>
<td>• AMSR (ADEOS II)</td>
<td>• AVHRR polar data (NOAA series)</td>
<td>• Digital Elevation Models (DEM$s$)</td>
</tr>
<tr>
<td>• SMMR (Nimbus 7)</td>
<td></td>
<td>• IceBridge</td>
</tr>
<tr>
<td>• SSM/I, SSMIS (DMSP series)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yearly Ingest</th>
<th>Total Archive</th>
<th>Yearly Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.7 TB</td>
<td>98.8 TB</td>
<td>150.0 TB</td>
</tr>
<tr>
<td>11.6 million files</td>
<td>27.2 million files</td>
<td>17.5 million files</td>
</tr>
</tbody>
</table>

* Metrics for 2010 calendar year
NSIDC User Services Office

Expert and timely assistance in selecting and obtaining data

Expertise in remote sensing, data manipulation and programming, outreach, and customer service

Support for data set documentation, data specific tools, and problem solving

Highest customer service satisfaction rating in the federal government (per external NASA reviews)
Green data center project

• Largely funded by a $500,000 NSF Grant
• Phase I complete: A 70% to 97% reduction in cooling energy while still meeting all hardware temperature and humidity specifications
• Phase II underway: Rooftop 50 KW solar array, expanded UPS
• Recipient of 2011 CO-LABS Governor’s Award for High Impact Research
Evolving as a data center

Data search and services: EDB and Polaris
Data and service casting: Libre
Interoperability: ACADIS

The above fall broadly under Informatics: “Delivering the right information, to the right person in the right place and time, in the right way”

Plotting our future:
Software architect
Linking with CU library system (Data Conservancy)

WMS map server

Google Earth
Enterprise Database (EDB) and Polaris

- EDB is the current major data management infrastructure project at NSIDC.
- EDB will serve as the authoritative, consistent, and extensible representation of NSIDC data holdings.
- Polaris leverages EDB’s file level metadata to provide subsetting, re-projection and other services.
- EDB replaces a collection of historical databases and utilizes PostGIS and the ISO 19115 metadata standards.

EDB is the infrastructure
Polaris is the interface
ACADIS: Advanced Cooperative Arctic Data and Information Service

ACADIS, a joint NSIDC/UCAR effort, is managing all Arctic data for NSF

Key year 1 goals:

• Continued data Ingest through CADIS as ACADIS system develops
• Form Data Advisory Committee
• Make all NSF Arctic data visible through NSIDC
• Prototype integrated/value added products
NSF: Exchange for Local Observations and Knowledge in the Arctic (ELOKA)

Why
- Local and community based knowledge of the Arctic is informing science, policy, and development.
- Communities look for their work to have influence beyond a particular research project and to be passed down to youth and shared broadly but ethically.
- There is an increasing need to manage the information so that local knowledge-holders can decide how to manage their “data” and how to share it with others effectively.

What
- ELOKA provides data management services and user support to facilitate the collection, use, exchange, and preservation, of local observations and knowledge.
- Currently have interviews, maps, and community measurements of sea ice, narwhal behavior, environmental change, and ecology.

ELOKA will become an arm of ACADIS
NSIDC will sustain the basic principles of data management

Preservation without access is pointless; access without preservation is impossible

Keep it simple and flexible: it’s about data, not systems!

Think about long-term archiving when planning for data collection

Document uncertainty

Involve scientists and users in data management, and data managers in science
Is a portal a one stop shop?

These ideas from the inspired mind of Mark Parsons
These ideas from the inspired mind of Mark Parsons
Some Bazaar Technologies

- Open metadata harvesting (OAI-PMH, CSW, etc.)
- Data Casting through RSS and Atom feeds to advertise data and any updates
- Service Casting to advertise related data services
- “Badging” data as open in a machine readable way
- Specialized aggregation of data casts and service casts
- OpenSearch

These ideas from the inspired mind of Mark Parsons
What if....

Finding data matching your interests was as easy as subscribing to the news?
What if....

Finding data matching your interests was as easy as subscribing to the news?

myData News.org

- Greenland 1 km DEM has been published
  A Digital Elevation Model (DEM) of Greenland acquired by A. Researcher is available in binary format at a 1 KM grid spacing in a polar stereographic projection ...
  more

- Greenland Ice Sheet Melt Characteristics Data updated

- Greenland Ice Sheet Melt Characteristics now available via OpenSearch API

Presented at the Data Stewardship: Data Discovery Session
27th Conference on Interactive Information Processing Systems (IIPS)
91st American Meteorological Society Annual Meeting
Libre - Aggregation Support

Search API
Use search API's to find casts for specific subject areas

Cast Crawler
• Monitor known casts for updates
• Find new casts on known sites

Seed list

Seed crawler
with new sites

Schedule

Heritrix

Feed the aggregator with new or updated casts

Aggregation Server
Support storage & retrieval of active cast entries

Retrieve a feed matching OpenSearch Criteria

Feed the aggregator with new sites

http://nsidc.org/libre/

API = Application Programming Interface
CC Zero waiver + norms

waive rights → public domain
+
attribution / citation through community norms, not a contract


CC = Creative Commons
Protocol for Implementing Open Access Data

Status of this Memo

This memo provides information for the Internet community interested in distributing data or

1. Intellectual foundation for the protocol

The motivation behind this memorandum is interoperability of scientific data.

The volume of scientific data, and the interconnectedness of the systems under study, makes integration of data a necessity. For example, life scientists must integrate data from across biology and chemistry to comprehend disease and discover cures, and climate change scientists must integrate data from wildly diverse disciplines to understand our current state and predict the impact of new policies.

The motivation behind this memorandum is interoperability of scientific data.

The volume of scientific data, and the interconnectedness of the systems under study, makes integration of data a necessity. For example, life scientists must integrate data from across biology and chemistry to comprehend disease and discover cures, and climate change scientists must integrate data from wildly diverse disciplines to understand our current state and predict the impact of new policies.

The technical challenge of such integration is significant, although emerging technologies appear to be helping. But the forest of terms and conditions around data make integration difficult to legally perform in many cases. One approach might be to develop and recommend a single license: any data with this license can be integrated with any other data under this
A Conceptual Architecture for the PIC