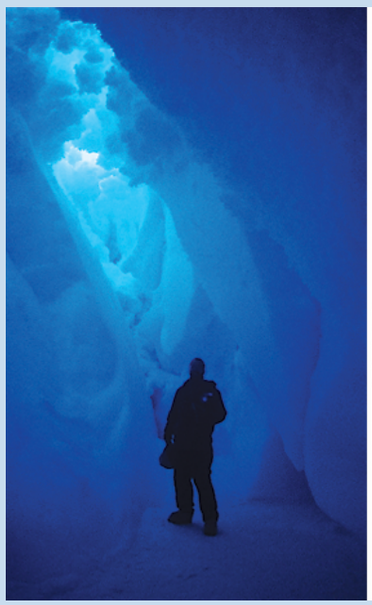
National Snow and Ice Data Center

Supporting Cryospheric Research Since 1976



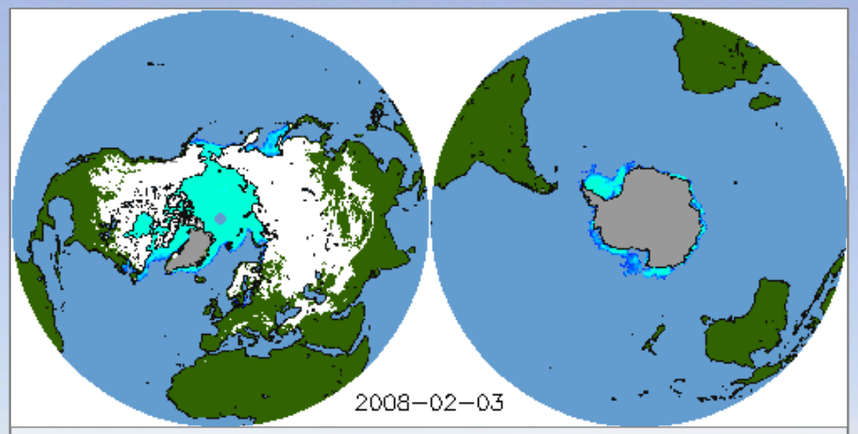
Data Management Planning at NSIDC

R. Duerr

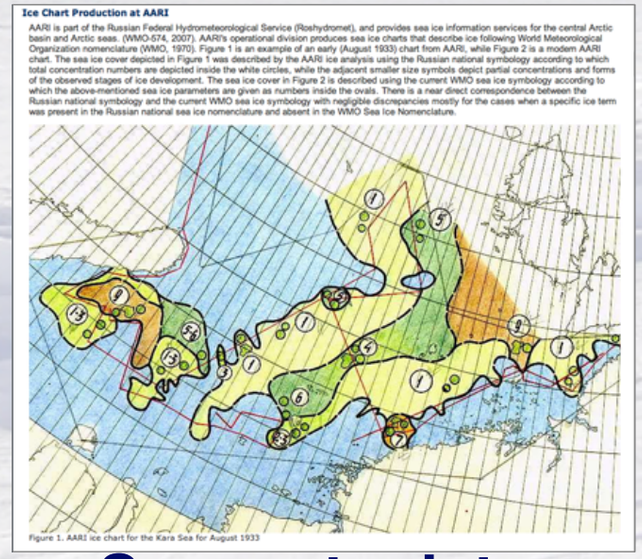
Outline

- **NSIDC background**
- **Data management planning at NSIDC**
- **Summary**
  - **What works**
  - **What needs work**

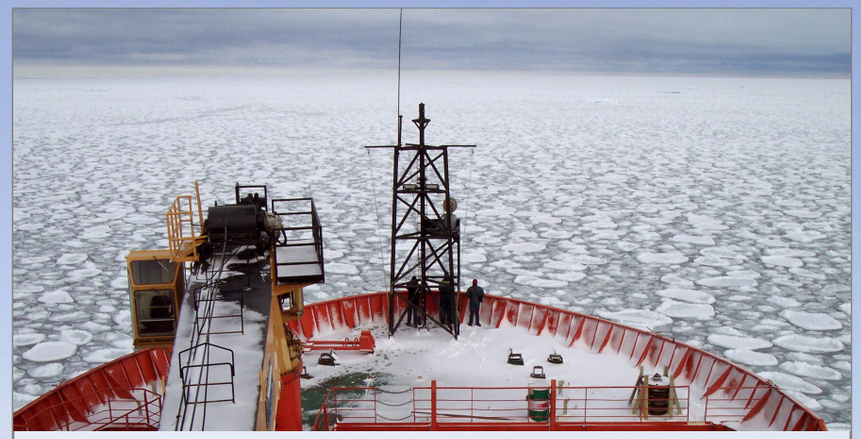
The National Snow and Ice Data Center…



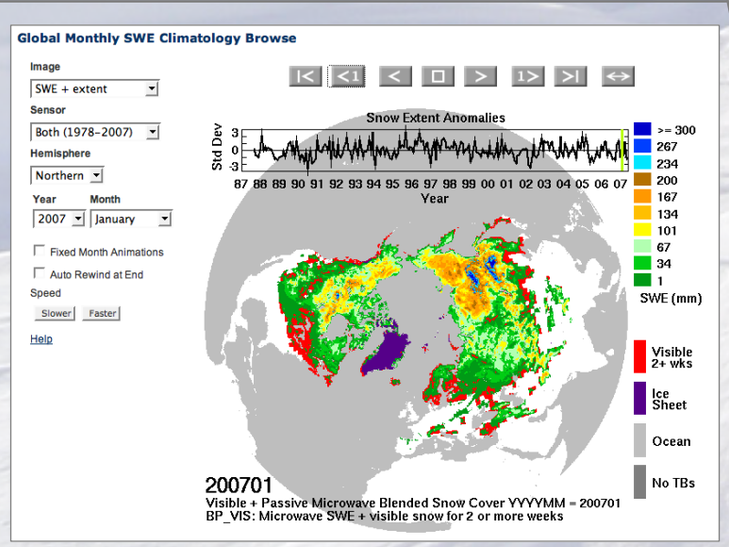
Manages and distributes scientific data



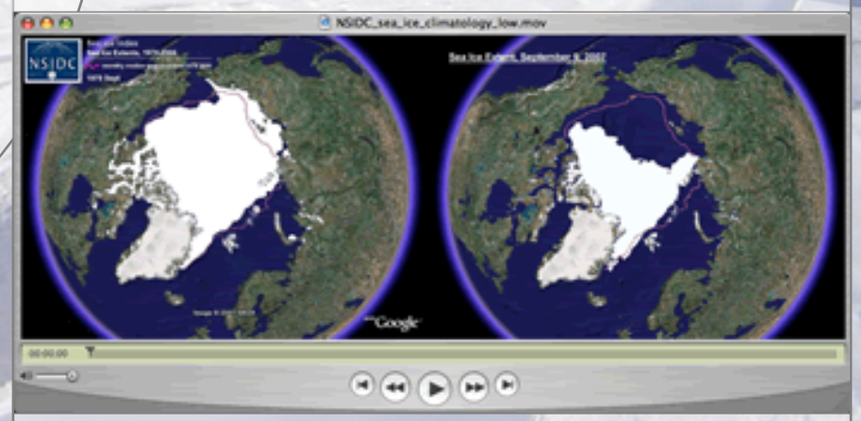
Supports data users



Performs scientific research



Creates tools for data access



Educates the public about the cryosphere

Focus on the cryospheric

Data…

Snow
Sea and lake ice
Glaciers
Ice sheets
Frozen ground
Ice cores
Icebergs
Avalanches
Elevation
Polar atmosphere
Radiation
Meteorology
Local and traditional knowledge

Used for…

Climate indicators
Climate modeling
Global energy balance
Water resources
Sea level change
Hazards
Education

Affiliations and sponsorship

Cooperative Institute for Research in Environmental Sciences

CIRES

University of Colorado at Boulder

World Data Center for Glaciology (since 1976)

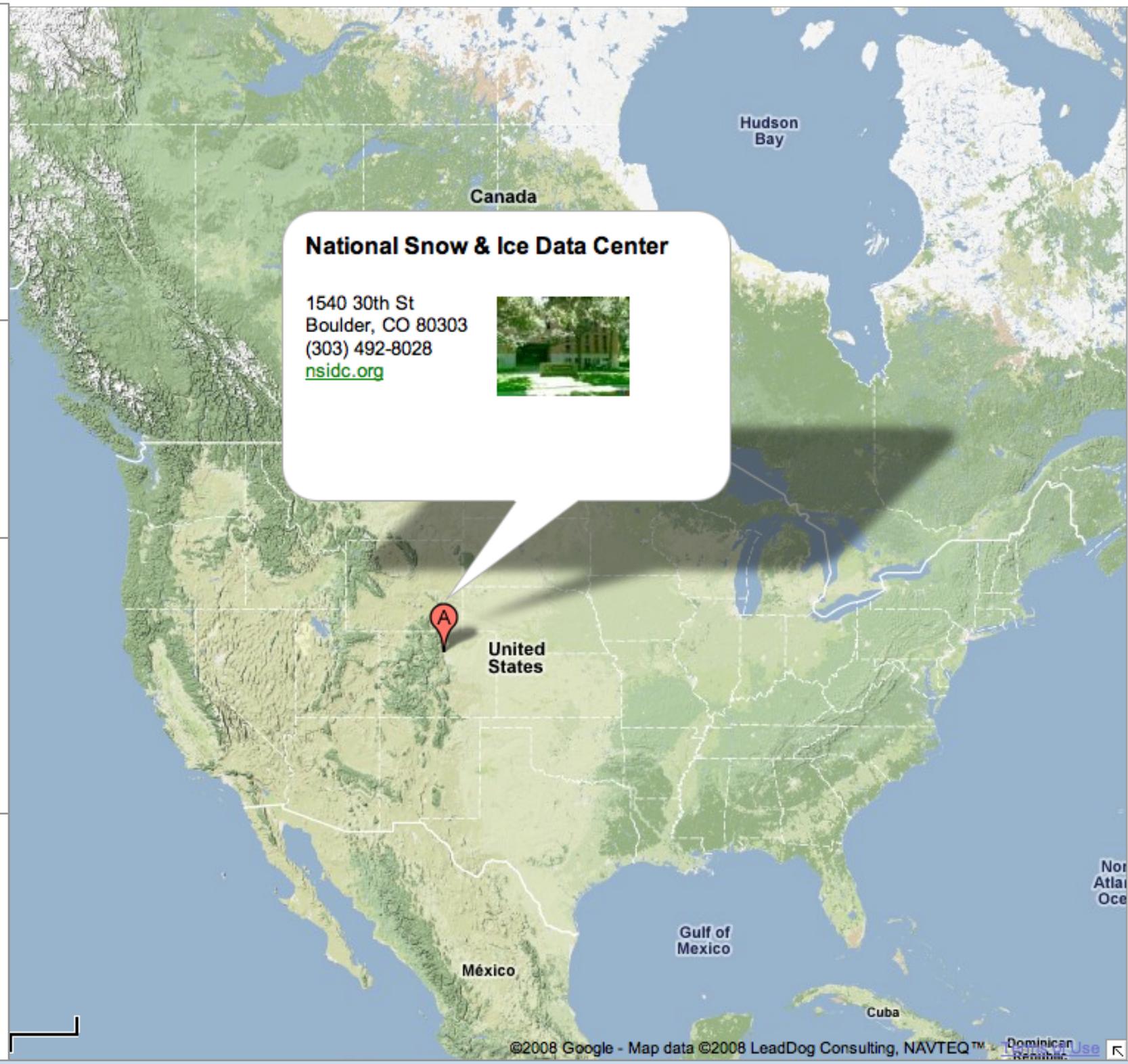
World Data Center for Glaciology Boulder

Main sponsors:

NSF

NASA

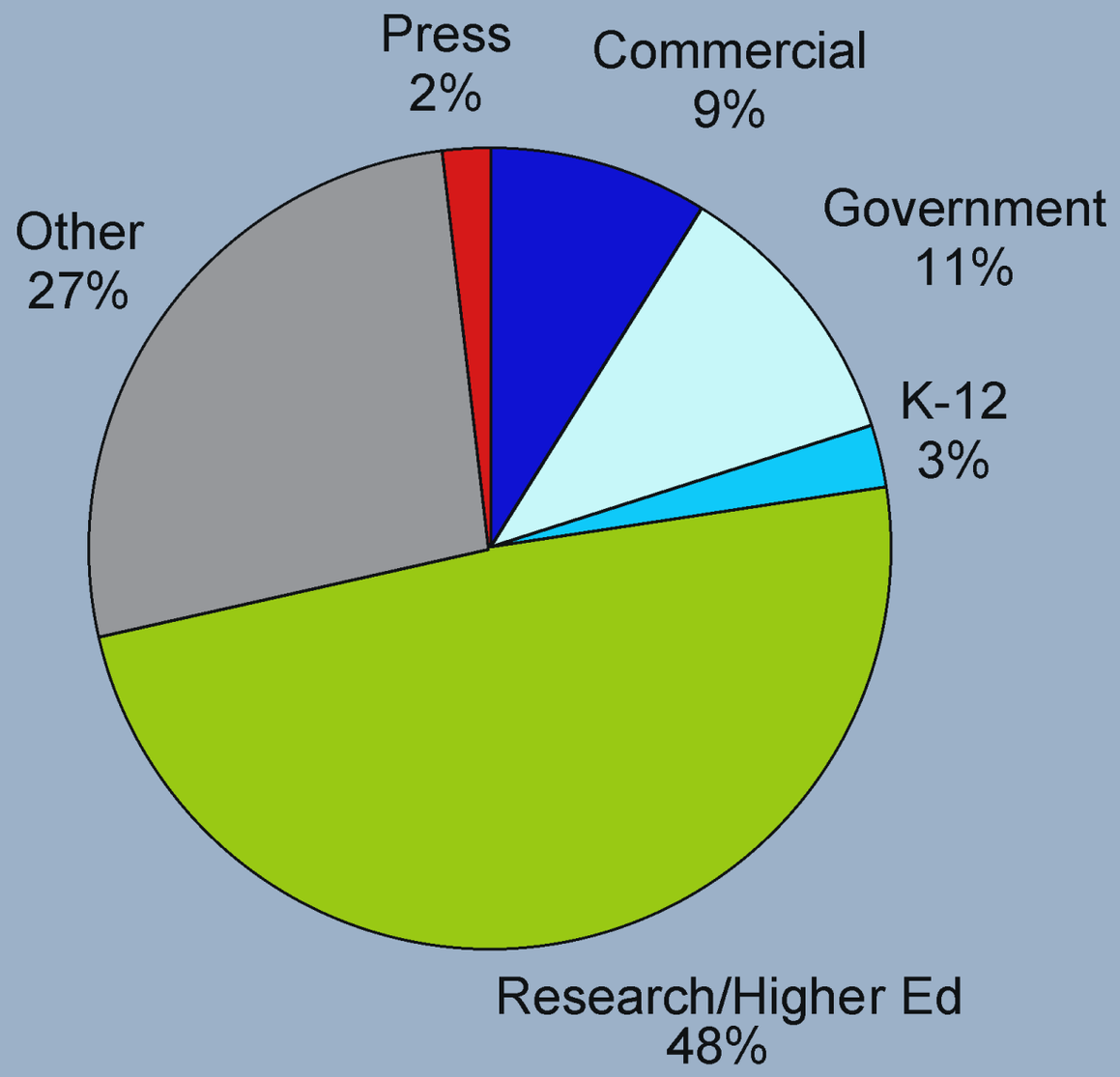
NOAA National Oceanic and Atmospheric Administration U.S. Department of Commerce



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

Agency-supported programs at NSIDC

NOAA at NSIDC

Data Conservancy

A Blueprint for Research Libraries

NSIDC DAAC

International Polar Year Data and Information Service

NSIDC

National Snow and Ice Data Center

University of Colorado at Boulder

photo: NOAA

Antarctic Glaciological Data Center

Frozen Ground Data Center

Cooperative Arctic Data and Information Service of the ARCTIC OBSERVING NETWORK

CADIS

GLIMS Glacier Database

Current foci at NSIDC

Cryospheric Research

Changes in the cryosphere – Arctic sea ice, mountain glaciers, Antarctic ice shelves, permafrost

Heat and water balances of the Arctic

Model parameterizations

Impacts on and responses of indigenous peoples

Data Management Research & Development

Data Conservancy – one of the first 3 nodes

Geospatial data

Discovery, visualization, and access

Long term archive (identifiers, citations, formats, interoperability)

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

Photo © Bjørn Anders Nyomen

Data Management Planning at NSIDC

- Varies dramatically by program, scale of project, and source of data
  - Planning for investigator data depends on the data management savvy/will of the PI
  - Planning for large programs may take place over years both before and after the program
  - Planning for small project data typically happens on submittal

What Works Well

- Levels of service

How Levels of Service Work

- LoS are a relative weighting of effort to achieve a certain state
- Five categories of services have been defined each with several possible desired states organized in order of increasing effort
- For each category a numeric score is given to each state
- Scores can be affected by a number of multiplicative factors, for example:
  - Data provided on obsolete media - multiply by 5
  - No format/structure modifications are needed - multiply by .75
- The adjusted scores for each category are summed to produce a single number for a data set
- The difference in scores between two Levels of Service reflect the effort required/freed by moving a data set from one LoS to another

Service Areas

- **Archival** – Levels of service in this area reflects the relative amounts of work required in order to ingest and archive a data set.

- **Metadata** – Levels of service in this area reflect the amount of work required to develop and maintain metadata not just for the data set as a whole; but also for any data element or service associated with the data set.

- **Documentation** – Levels of service in this area reflect the amount of work required to document the data set and any associated web pages.

- **Distribution** – Levels of service in this area reflect the amount of work required to support data distribution or distribution-related services.

- **USO Support** – Levels of service in this area reflect the amount of work required to support human-human requests for information about or help with a data set (via any mechanism – phone, email, etc.)

What Works Well

- Levels of service
- Involving data managers in project planning

An Example - The CLP Experience

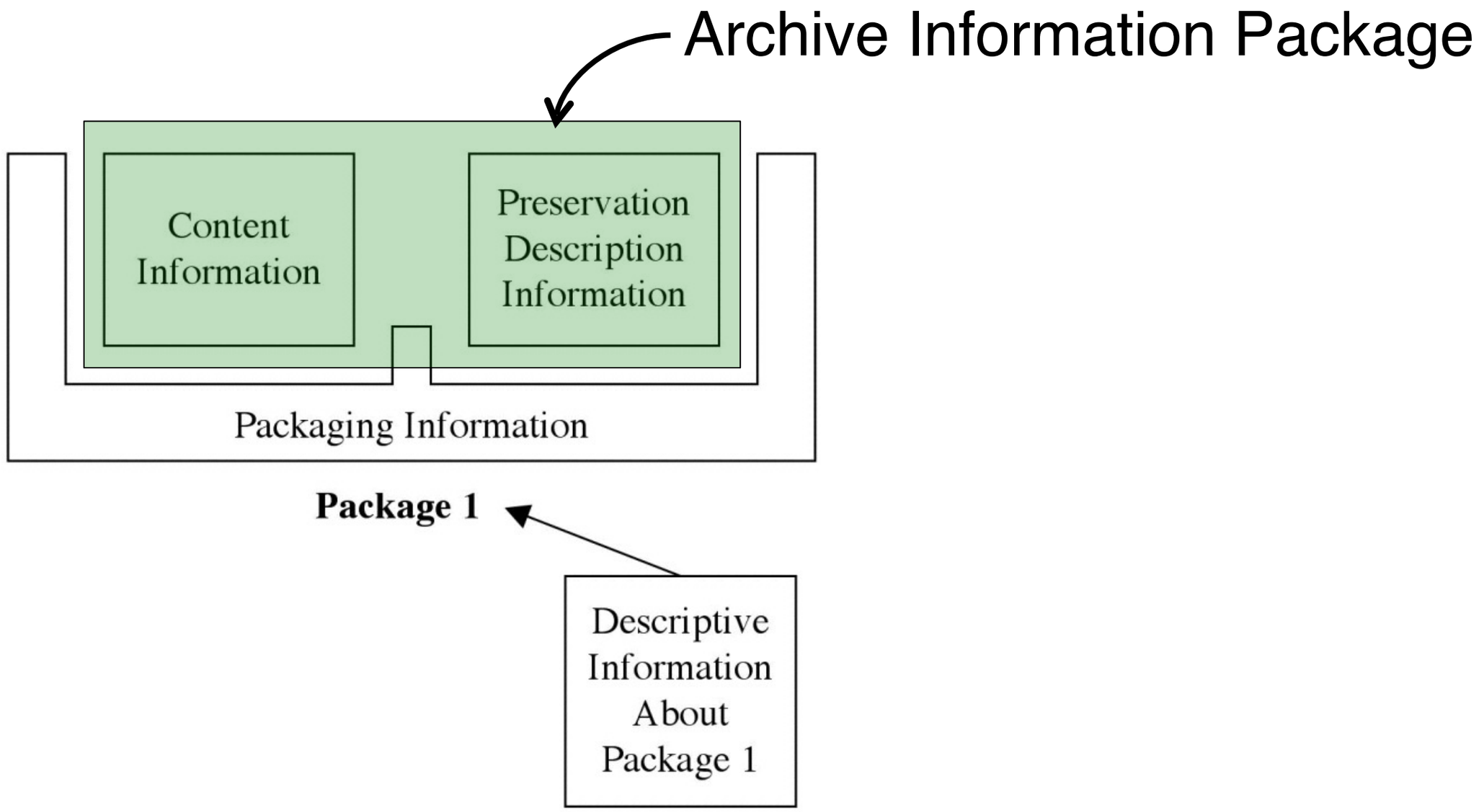
- NSIDC was involved from the start
- NSIDC management folks were in the field
  - Interviews and follow up with the investigators
  - Manual and automated QC of the data collected each day
- Resulted in better QC documentation and higher-quality data

What Doesn't

- Trying to provide the same level of service for all data sets

- System handling of provenance, context, etc.

OAIS Reference Model[1]



[1] Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, Blue Book, January 2002.

OAIS Reference Model

- Content Information
  - The data object to be preserved
  - Information that describes the data object
    - Typically interpreted as the syntax and semantics of the file structure
- Preservation Description Information
  - Provenance – Origin or source of the data, any changes that have taken place since, and who has had custody of it
  - Fixity – the authentication mechanisms (with keys) needed to ensure that the data object has not been altered in an undocumented manner
  - Reference – identification mechanisms and values
  - Context – relation of the object to its environment

Contextual Information for the Earth Sciences:

- **Instrument/sensor characteristics** including pre-flight or pre-operational performance measurements (e.g., spectral response, noise characteristics, etc.)

- **Instrument/sensor calibration data** and method

- **Processing algorithms** and their scientific basis, including complete description of any sampling or mapping algorithm used in creation of the product (e.g., contained in peer-reviewed papers, in some cases supplemented by thematic information introducing the data set or derived product)

- Complete information on any **ancillary data** or other data sets used in generation or calibration of the data set or derived product

Contextual Information (continued):

- **Processing history** including versions of processing **source code** corresponding to versions of the data set or derived product held in the archive
- **Quality assessment** information
- **Validation record**, including identification of **validation data**
- In the case of earth based data, station location and any changes in location, instrumentation, controlling agency, surrounding land use and **other factors** which could influence the long-term record

Contextual Information (continued):

- A **bibliography** of pertinent Technical Notes and articles, including refereed publications reporting on research using the data set
- **Information received back from users** of the data set or product

What Doesn't

- Trying to provide the same level of service for all data sets
- System handling of provenance, context, etc.
- Trying to find one size fits all solutions

Science World Views

- To the GIS community, the world is:
  - A collection of features (e.g., roads, lakes, plots of land) with geographic footprints on the Earth (surface)
  - The features are discrete and are described by a set of (typically 2-D) characterstics such as shape/geometry

- To fluid-earth scientists, the world is:
  - A set of observations/measurements described by parameters (e.g., temperature, velocity) that vary as continuous functions in 4-D space-time
  - Parameter behaviours are governed by a set of equations

borrowing from Ben Domenico and Stefano Nativi

Closing thoughts