SEARCH Data Management Strategy

Contributors

- Florence Fetterer (NOAA-NSIDC)
- Tom Marr (UAF-IAB)
- A. David McGuire (UAF-IAB-USGC)
- Jim Moore (NCAR)
- Mark Parsons (NSIDC)
- Mohan Ramamurthy (UCAR-Unidata)
- Ron Weaver (NSIDC)
- Steve Worley (NCAR).

Outline

- The Role of Data Management in SEARCH

- Design Issues and Options

- Recommendations from SIW

- Issues Requiring Immediate Attention
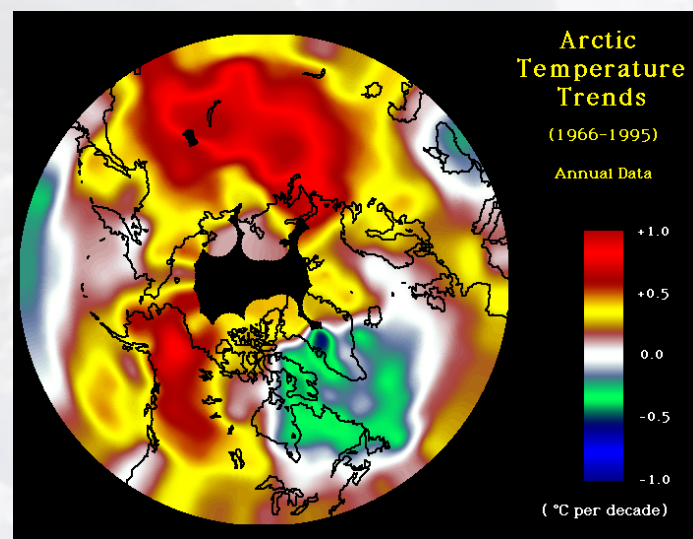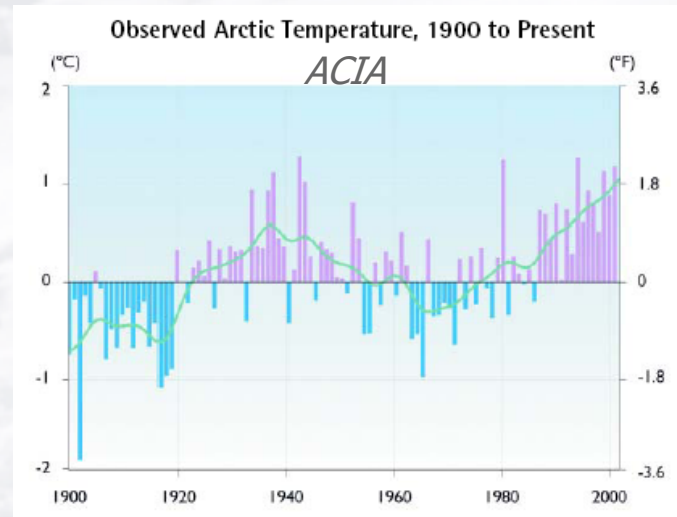
- Data Management Priorities

The Role of Data Management in SEARCH

- Design and implementation must effectively and efficiently support SEARCH science, education, and outreach

- Should build upon disciplinary expertise in data management

- Methods should be flexible and extensible, and use community standards and common "best practices"

- Should provide
  - Data discovery across disciplines and program
  - Open data distribution
  - Data archival and stewardship

Definition of Data Collections

- **Three Categories**
  - Research - PI
  - Community – network
    Especially relevant for SEARCH to support change detection, attribution, and prediction studies
  - Reference - NWS

John Walsh et al.

Design Issues and Options

- Need a distributed data management system
  - capitalizes on existing expertise across disciplines
- Need data and information coordination service
  - central data portal
  - should have an advisory group
- Need to plan for integration
  - to support multi-disciplinary synthesis
  - can be full or partial, and implemented gradually depending on needs of the program, resources available, and presentation of new opportunities
  - central portal is essential, but can be simple or complex (more complex requires more metadata)
- Need to coordinate with other data management activities
  - IPY, ISAC, CliC
  - Policies and practices need to coordinated with above activities and with other organizations (IASC, ICSU, AOSB)

Design Issues and Options (cont.)

- Need for free and open access to support science, education, and outreach
- Unique needs of social data
  - Data transfer methods/interfaces/agreements that define rights and responsibilities of data archives, data providers, and data users, particularly issues of proprietorship, fair use, and privacy
  - Appropriate data access that maintains orginal content of data while maintaining privacy
  - Appropriate data attribution communities/individuals need to maintain ownership of their knowledge and be recognized for their contributions
  - Data description (metadata) to facilitate data discovery and integration
  - Need to preserve myriad data types including text, maps, video, audio, imagery, databases, and narrative documents
  - Compliance with relevant policies for research involving human subjects (AFN Guidelines for Research, Federal Policy for Protection of Human Subjects, and corresponding regulations in other countries
- Need to support education and outreach
  - Opportunity and challenge for an end-to-end service (science, education, and outreach)
  - Need for processing before being used as educational materials
  - Digital Library for Earth Science Education, which is evolving, may be model

Recommendations

- Form a SEARCH Data Management Advisory Group
  - To develop comprehensive data management plan and data policy to advice data coordination service
  - SDMAG: members from panels, IPMC, SSC, and 3 to 4 data managers
- SDMAG should develop a SEARCH data policy
  - To address issues of data accessibility, timeliness, attribution, and security
- Develop data and information coordination service
- Data, metadata, interfaces, and tools should adhere to existing community standards
- SEARCH should encourage the rescue and incorporation of relevant historical data
- Agencies participating in SEARCH need to support long-term stewardship of SEARCH data

Recommendations (cont.)

- Announcements of Opportunity should include the following data management requirements for investigators
  - Investigators should provide a metadata inventory description within 3 months of being funded to the data coordination service and relevant archive
  - Investigators must specify in their proposal where the data will be archived
  - Investigators must specify in their proposal a point of contact for data submission and documentation
  - Investigators must specify in their proposal which data are destined to be "community" data, which are to made available quickly after collection/verification
  - Every project must submit complete documentation and quality controlled data in accordance with SEARCH data policy
  - Investigators submitting proposals must adhere to the general policy of the agency to which they are proposing as well as to the specific terms of the SEARCH data policy

Issues Requiring Immediate Attention

- AOs should include recommendation 7
- SDMAG should be formed
- SDMAG should develop SEARCH data policy
- SDMAG should facilitate development of SEARCH data inventory, based on needs of the panels
  - Early recognition of data gaps and inconsistencies in the proposed measurement strategies
- SDMAG should work with IPMC and SSC to foster cooperation with international projects and organizations, especially in preparation for IPY

SEARCH Data Management Priorities

- Form SEARCH Data Management Advisory Group (highest priority)

- Develop SEARCH Data Policy (highest priority)

- Develop comprehensive SEARCH Data Management Plan (highest priority)

- Develop SEARCH Data Inventory (highest priority)

- Establish SEARCH data management requirements for investigators (highest priority)

- Create a central SEARCH Data and Information Coordination Service (highest priority)

- Rescue and incorporate relevant historical data (intermediate priority)

Data Working Group Draft Terms of Reference

Advises and aids the IPMC and the SSC in SEARCH Data Management to provide free and open access to data in support of SEARCH science, broad community and lay users, education, and outreach through the following activities:

- Development of SEARCH Data Policy

- Development of comprehensive SEARCH Data Management Plan

- Establishment of SEARCH data management requirements for investigators

- Provide a conduit for establishing broad community and lay user requirements and meeting their needs for data

- Provide guidance and advice to the central SEARCH Data and Information Coordination Service

- Work with IPMC and SSC to foster data management cooperation with international projects and organizations, especially in preparation for IPY

Data Working Group Members

David McGuire (Acting Chair – SEARCH SSC)

Florence Fetterer (NSIDC)

Larry Hamilton (SEARCH Responding Panel)

Jim Moore (JOSS-UCAR)

Mark Parsons (NSIDC)

Mark Serreze (SEARCH Understanding Panel)

Mike Steele (SEARCH Observing Panel)

Steve Worley (UCAR)

Data Working Group Draft Terms of Reference

Advises and aids the IPMC and the SSC in SEARCH Data Management to provide free and open access to data in support of SEARCH science, broad community and lay users, education, and outreach through the following activities:

- Development of SEARCH Data Policy

- Development of comprehensive SEARCH Data Management Plan

- Establishment of SEARCH data management requirements for investigators

- Provide a conduit for establishing broad community and lay user requirements and meeting their needs for data

- Provide guidance and advice to the central SEARCH Data and Information Coordination Service

- Work with IPMC and SSC to foster data management cooperation with international projects and organizations, especially in preparation for IPY

SEARCH Data Policy

The SEARCH Data Policy applies to all data defined as SEARCH data by the participating agencies. An initial list is contained in the SEARCH Implementation Workshop Report. The main purpose of the SEARCH data policy is to maximize data access, integration, and ultimately long-term preservation. It is intended to complement policies from other associated programs such as International Polar Year (IPY), Arctic Observing Network (AON), and ARCSS Synthesis of Arctic System Science (SASS). The participating agencies in SEARCH should enforce the policy as described below. When SEARCH data are collected as part of a collaborative international effort they should also be distributed according to SEARCH data policy of free and open access.

SEARCH Data Policy (cont.)

SEARCH data must be made fully, freely, and openly available as soon as possible after collection and quality control. Investigators must describe their plan and timing for making data available as part of their proposals. Timely data availability will be part of SEARCH proposal review criteria.

In keeping with the IPY data policy, the only exceptions to this policy are:

- where human subjects are involved, confidentiality must be protected
- where local and traditional knowledge is concerned, rights of the knowledge holders shall not be compromised
- where data release may cause harm, specific aspects of the data may need to be kept protected (for example, locations of nests of endangered birds or locations of sacred sites).

SEARCH Data Policy (cont.)

Data policy actions for SEARCH PIs are:

Make all SEARCH community project data available as quickly as possible subject to procedures approved in the proposal.

Follow guidelines for the preparation and submission of data, metadata, and documentation as described in the SEARCH Data Management Plan. Standard metadata are required to achieve the program goals of maximum data integration and synthesis, broad community usage, and long-term preservation. Make sure all project related data sets are submitted to an appropriate data archive. These archives are generally defined as national archive centers and/or SEARCH project specific repositories.

SEARCH Data Policy (cont.)

All SEARCH data users are obligated to make proper attribution of the data providers. Attribution should credit both the data provider or author and the data center or publisher. In a scientific publication, attribution should take the form of a formal citation, such as for a journal article or book. General acknowledgement may be more appropriate in other situations such as instrument development, field applications, or logistics planning.

Data Working Group Draft Terms of Reference

Advises and aids the IPMC and the SSC in SEARCH Data Management to provide free and open access to data in support of SEARCH science, broad community and lay users, education, and outreach through the following activities:

- Development of SEARCH Data Policy

- Development of comprehensive SEARCH Data Management Plan

- Establishment of SEARCH data management requirements for investigators

- Provide a conduit for establishing broad community and lay user requirements and meeting their needs for data

- Provide guidance and advice to the central SEARCH Data and Information Coordination Service

- Work with IPMC and SSC to foster data management cooperation with international projects and organizations, especially in preparation for IPY