Integrated Data Management for ARCSS Projects

JOSS Breaks the Ice on Field Data Archival and Exchange

This article was prepared by James A. Moore, Gregory J. Stossmeister, and Steven D. Roberts, all of the Joint Office for Science Support, University Corporation for Atmospheric Research.

The University Corporation for Atmospheric Research (UCAR) Joint Office for Science Support (JOSS) provides scientific, technical, and administrative support to the scientific community in planning, organizing, and implementing research programs associated with field projects worldwide. The National Science Foundation and the National Oceanic and Atmospheric Administration are its principal sponsors.

Access to and integration of multidisciplinary data from field projects recently completed or underway is critical to the timely and accurate understanding of the rapid changes that are now occurring in the Arctic. The NSF Arctic System Science (ARCSS) program is committed to facilitating data archival and providing easy mechanisms for data exchange among researchers interested in the Arctic system. JOSS offers some specific capabilities that address these two important ARCSS objectives.

JOSS has been involved in data management support for a number of ARCSS field projects, both domestic and international, including the Surface Heat Budget of the Arctic Ocean (SHEBA), Arctic Transitions in the Land–Atmosphere System (ATLAS), the International Tundra Experiment (ITEX), the Western Arctic Shelf–Basin Interactions (SBI) project, and the Arctic Regional Climate Modeling Intercomparison Project (ARCMIP). JOSS also participates on committees to further improve the collection, archival, and dissemination of all manner of Arctic data sets.

This paper discusses the organization of JOSS in the context of field project support and related data management activities, describes data management strategies that have been successfully implemented in ARCSS projects, provides some examples of specific support to projects, and discusses some of what the future holds for data management support to ARCSS field research.

Overview of JOSS Support Services

JOSS has a staff of skilled and experienced scientific, technical, and administrative specialists who collaborate extensively with geophysical scientists and organizations to assist them in planning, organizing, and conducting research by supporting scientific planning efforts, collaborative research programs, field experiments, and data management. Collectively the JOSS staff has decades of experience at these tasks, and the office has supported programs of all sizes worldwide for more than 20 years. JOSS adds value to the scientific endeavor through an integrated approach to the full life of projects (science, technology, data management, logistics, and administration).

JOSS Data Services

JOSS has worked for over 15 years to develop its data management support capabilities, which have contributed to the success of research projects in many locations and disciplines. Services include collecting complete, high-resolution, high-quality data sets, supporting project objectives, and providing tools to view these data during both the field and analysis phases of a project. One such tool is the JOSS field catalog, which has proven particularly useful for distributing preliminary data sets in the field, providing access to data products needed for operational decision making, and maintaining a running assessment of project accomplishments.

JOSS has developed and maintains a state-of-the-art data archive and dissemination system that provides single-source access to complete project data sets. Most data are accessible for browsing and ordering online, with connectivity to distributed archives. Early versions of data sets can be shared with limited access (under password pro-

tection) to expedite the timely exchange and integration of related measurements. Data are delivered at no charge over the Internet or via CD-ROMs or other media.

Capabilities are available to provide value-added data processing and quality control procedures to ensure the best possible research-quality data sets. A range of data processing, quality control, and documentation procedures is available, including format conversions, automated and visual data consistency checks, confirmation of uniform metadata, and formation of data composites.

The details of the data system will be described in the context of how it supports the Arctic researcher. This includes the ability to display and download data sets online, extract parameters from different data sets and create new composite data sets, and prepare and distribute project CDs. These capabilities are critical for supporting a number of Arctic regional data and model intercomparison efforts now underway or planned, as well as outreach activities to better inform the public of ARCSS research.

Field Project Support

JOSS is organized to help investigators with all facets of field project support, including guiding and assisting in developing sampling strategies, implementing support services such as an operations center or field catalog, and directly supporting field operations, including operations coordination and field data management services. The support may include organizing and leading site selection and a broad range of site arrangements including site preparation and modifications, contracting, workspace and laboratory infrastructure, communications, and information management. Project logistics including shipping and travel arrangements can be provided. JOSS has broad experience coordinating multi-agency and multinational facilities (aircraft, ships, and ground-based) at both domestic and remote foreign sites in field campaigns led by both U.S. and international scientists from numerous scientific disciplines.

Data Management Support to ARCSS Field Projects

The following description of support provided to NSF ARCSS field projects is based on 7 years of dedicated support to ARCSS-funded projects and 20 years of experience supporting multidisciplinary field projects around the world. The five ARCSS projects are:

- Surface Heat Budget of the Arctic Ocean (SHEBA), a multiphase international project to investigate the role of Arctic climate in global change;
- Arctic Transitions in the Land–Atmosphere System (ATLAS), a coordinated program that will examine the geographical patterns and controls over climate–land surface exchange and develop reasonable scenarios of future change in the Arctic;
- International Tundra Experiment (ITEX), a MAB NSN (Man and the Biosphere Northern Sciences Network) initiative established in 1990 to monitor the performance of plant species and communities on a circumpolar basis in undisturbed habitats with and without environmental manipulations;
- Western Arctic Shelf–Basin Interactions (SBI) project, a multiyear, interdisciplinary program to investigate the impact of global change on physical, biological, and geochemical processes over the Chukchi and Beaufort Sea shelf basin region in the western Arctic Ocean; and
- Arctic Regional Climate Modeling Intercomparison Project (ARCMIP), an international intercomparison of Arctic regional climate and mesoscale model simulations

JOSS data archives for ARCSS field projects. A total of 60.4 GB of data have been archived, processed, and made available to researchers.

Project Name	Years of Support	Volume of Data (GBytes)
SHEBA	1997–2003	49.7
ATLAS	1999–2003	2.7
ITEX	1999–2003	0.8
SBI	2001–2006	2.1
ARCMIP	2001–2004	5.1

A Workable Strategy

An integrated data management strategy is important to assure that complete data archives are provided to project scientists and the larger science community in a timely and efficient manner. Field projects benefit from the implementation of sound data management procedures and protocols before any data are collected. This includes the specification of a data policy, consideration of the data format, and development of documentation guidelines that maximize the ease of data exchange and archival. JOSS has implemented a data

management system to facilitate submission, archival, and distribution of project-related data sets.

JOSS works with the science management offices, project offices, and individual investigators to support ongoing projects while fostering a consistent data management strategy that makes sense for the project's science objectives. JOSS assists the project scientists in determining data management requirements for the field phases and associated analysis periods to maximize efficient and timely data exchange. JOSS also collaborates with the Arctic Data Coordination Center (ADCC) at the National Snow and Ice Data Center (NSIDC) in the project planning phase to provide guidance to project investigators on effective data management strategies. This includes the provision of initial metadata for anticipated data sets. The ADCC provides a permanent archive location for data from all ARCSS projects.

The Field Catalog

If the project includes field phases, JOSS can provide a web-based, online field catalog or project web pages to support near-real-time documentation of activities and selected data displays. This also includes interactive access to common data sets of interest and sharing of preliminary data and analyses among project scientists who are in the field and elsewhere. The field catalog allows automatic and/or semi-automatic submission of field reports and data products (such as satellite images or preliminary research products and plots) for review and exchange while the field project is underway. Operational summaries, instrument status reports, daily mission plans, and other specialized reports are also ingested into the catalog.

The JOSS field catalogs in ARCSS-supported projects have proven to be valuable for reporting and monitoring operational activities and as a permanent archive of field activities. The catalogs from each field deployment, including ship cruises and multiyear field deployments, are kept in the JOSS archive and can be accessed at any future time.

Archival of Supporting Data Sets

Some field projects may require the collection of supporting data sets that add to the richness of the complete project data archive. JOSS typically completes a search of relevant available data and establishes data quality. High-resolution operational data (such as data available from national, regional, or local agencies) are often not routinely archived. Examples of these types of data include satellite imagery, numerical model products, high-time-resolution surface data from national or regional networks, and complementary data from nearby research projects. JOSS often makes special arrangements to archive these data for later access by project scientists.

Data Archive and Distribution System

JOSS has developed and implemented a Data Management System (termed CODIAC) that offers scientists a means to submit their data and accompanying metadata, identify and download other data sets of interest, display selected data sets online, and update data sets and documentation as necessary during the life of the project. For ARCSS, JOSS acts as an interim archive for field project data sets. This system provides a rapid turn-around of preliminary and updated data sets and password protection during the initial analysis period.

Since JOSS operates as an interim archive for ARCSS data sets, it is important to maintain close coordination with permanent archive centers. JOSS works with ADCC to assure timely transfer of data and documentation in a way that minimizes disruption to the access of project-related data sets.

Special Data Processing

It is possible to provide specialized processing for selected data sets, including quality control of data sets, parameter extraction from different data sets, and data set compositing (combining parameters from different data files or merging multiple data files). In addition, JOSS works with data providers and investigators to maintain consistent data formats and documentation for the supported project.

Finally, JOSS provides support to project scientists in integrating their data sets for education and outreach. This is done through online access to data and publications from the project and the compilation of CD-ROMs that focus on activities at a single site or region of the field deployment.

Specific Examples of Customized Support

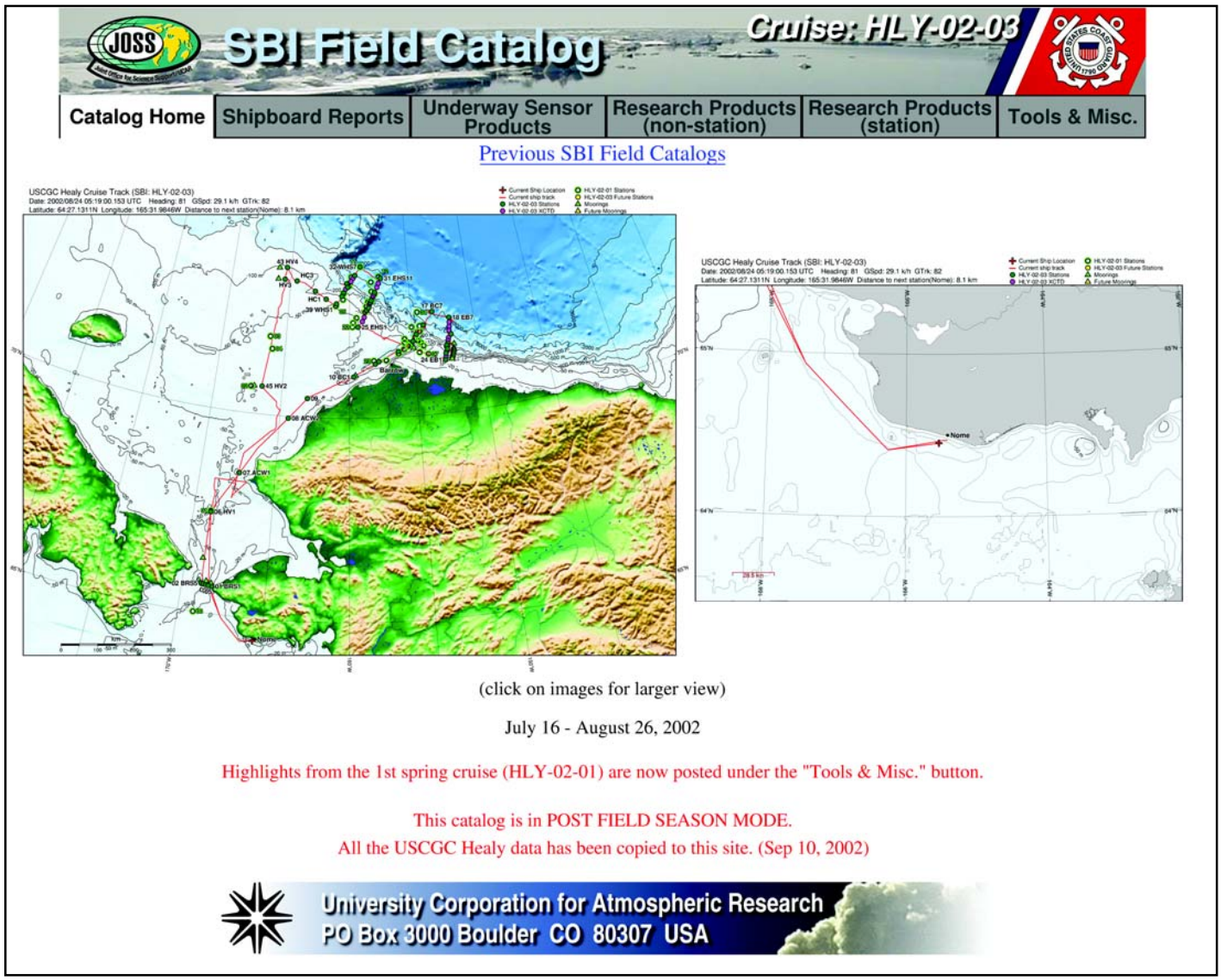
The general information presented above refers to services that JOSS provides to ARCSS projects as requested. The following examples show specific assistance that JOSS has given to projects.

The SBI Field Catalog was implemented by JOSS aboard the USCGC Healy *and is located at www.joss.ucar.edu/sbi. Navigation links access a variety of products that summarize the cruise operations. Products and reports are updated aboard ship as the cruise progresses. The larger regional image to the left shows the ship track with primary stations and mooring locations. The close-up image on the right provides details. Images are updated every 15–30 minutes.*

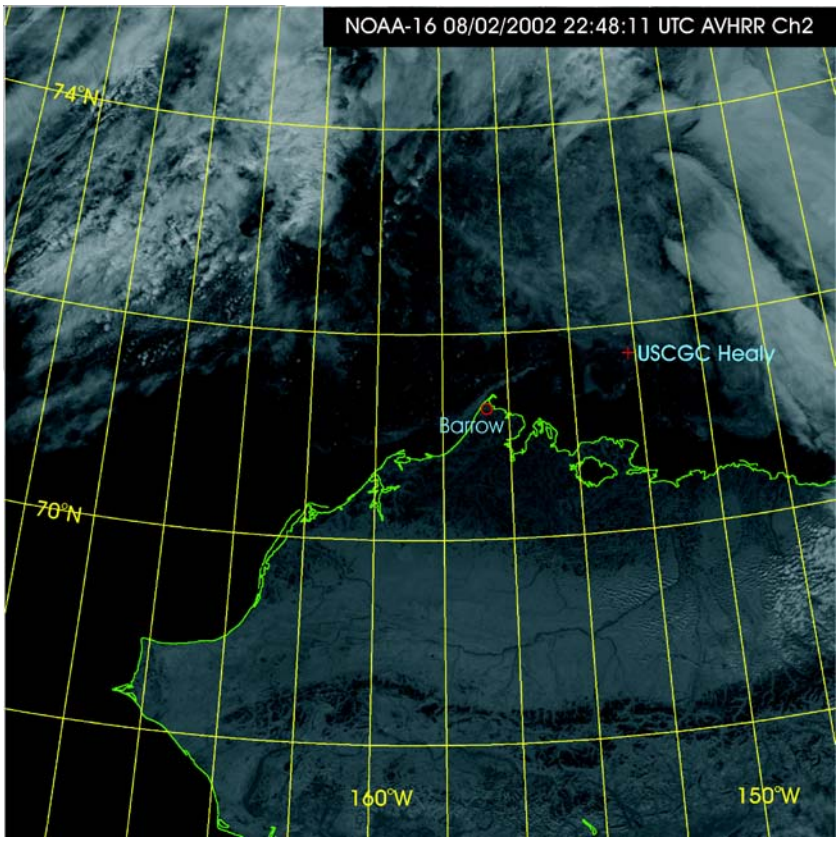
Support to SBI using the JOSS Field Data Catalog

During SBI, JOSS implemented an online field catalog during selected cruises to provide near-real-time documentation and browsing of operational data collected aboard ship. Previously deployed in the Arctic for the SHEBA project, the field catalog organized browse products and documentation for use in the field and provided a detailed field summary report for researchers after field operations ended. The catalog facilitated communication among researchers in the field and kept project participants abreast of ongoing operations. In addition, a portion of the shipboard catalog was routinely uplinked via satellite to the JOSS facilities in Boulder, Colorado. The mirror catalog in Boulder was used by project participants ashore, the USCG staff and families at home, the Arctic Eskimo Whaling Commission, and others interested in the ship's operations.

During SBI, two components of the field catalog were especially popular: a ship track plot updated every 15–30 minutes showing past and future station locations, moorings, and bathymetry (using data from the International Bathymetric Chart of the Arctic Ocean) relative to current ship position (see p. 27); and an event log detailing station activities, times, depths, and locations. In addition, during the spring cruise a form was implemented as part of the field catalog that allowed researchers to log detailed ice observations complete with digital photos and automatic reporting of current underway data from the ship, including position, water depth, and weather conditions.



AVHRR imagery of northern Alaska and nearby Chukchi Sea from NOAA-16. This image was collected aboard the USCGC Healy *during the summer SBI cruise of 2002. The position of the* Healy *at the time this image was taken is shown east northeast of Barrow. The image shows largely open water west of Alaska (black area), while sea ice of varying concentration exists in the Barrow vicinity and just offshore of the north coast. Significant cloud cover is seen in the northwestern and northeastern portions of the image and partially obscures the sea ice north of the* Healy.



Catalog Products

One of the important features of the catalog architecture is that it allows customization of products and displays to meet the needs of the participants. Before the SBI deployment aboard ship, JOSS staff worked with the onboard scientists to put together a list of operational and research products that would aid in data analysis during and after the cruise. These products included:

- Satellite products [each overpass of NOAA and DMSP (Defense Meteorological Satellite Program) polar orbiters]:
  - 0.5-km-resolution visible images from DMSP satellites over an approximately 400- × 400-km area centered on the ship,
  - 1-km-resolution visible images from both types of satellites over an 800- × 800-km area centered on the ship, and
  - 3-km-resolution visible and infrared images over an approximately 1800- × 1800-km area centered on 70°N, 165°W;
- Ship track (updated every 15–30 minutes):
  - regional and zoom maps with track, stations, moorings, and bathymetry;
- Weather and other data (updated twice daily):
  - 24-hour time series plots of temperature, winds, pressure, humidity, and water depth;
- Research data (not station specific; spring cruise only):
  - ice observation reports (as frequently as every 2 hours when underway),
  - digital pictures of ice conditions, and
  - wildlife photos;
- Research data (station specific):
  - CTD data: WHP (WOCE Hydrographic Program) exchange format compressed (zip) files; comments on each cast; CSV (comma-separated value) format ASCII data; standard plots of temperature, salinity, oxygen, transmittance, photosynthetically active radiation (PAR), and fluorometer measurements; and vertical section plots of various parameters for station transects, and
  - bottle data: WHP exchange format files, bottle hydrographic reports, and synthesized vertical section plots of various variables for various station transects;
- Event logs:
  - detailed station event logs for each station with time and location, a map of the station location, both in-water and on-deck times, and Seabeam water depths for each event, based on daily logs generated by the U.S. Coast Guard;
- Reports:
  - chief scientist's daily operational summaries,
  - Teacher at Sea (TEA) daily journal reports (summer cruise only),
  - cruise summary reports for each principal investigator,
  - service team cruise summary reports, and
  - service team cruise science reports.

Post-Cruise Catalog Use

As mentioned above, a significant subset of all field catalog reports, ship location maps, satellite imagery, and periodic ice observations (spring cruise), as well as all station data, were mirrored to the JOSS facility in Boulder on a daily basis so that land-based co-workers and other interested parties could monitor ship operations. Ship location information was mirrored with a 12-hour time delay at the request of the Coast Guard because of security considerations. Research-quality data were protected for the principal investigators by implementing password control on the appropriate portions of the Boulder catalog. (This password protection will remain in place until the data enter the public domain.)

Near the end of each cruise, JOSS personnel onboard the *Healy* worked with service team

members to produce a CD of the catalog station products, as well as all service data and other ancillary data sets produced during the cruise. These CDs were distributed at the end of each cruise to all interested parties, along with a CD copy of the ship's underway data files produced by the U.S. Coast Guard.

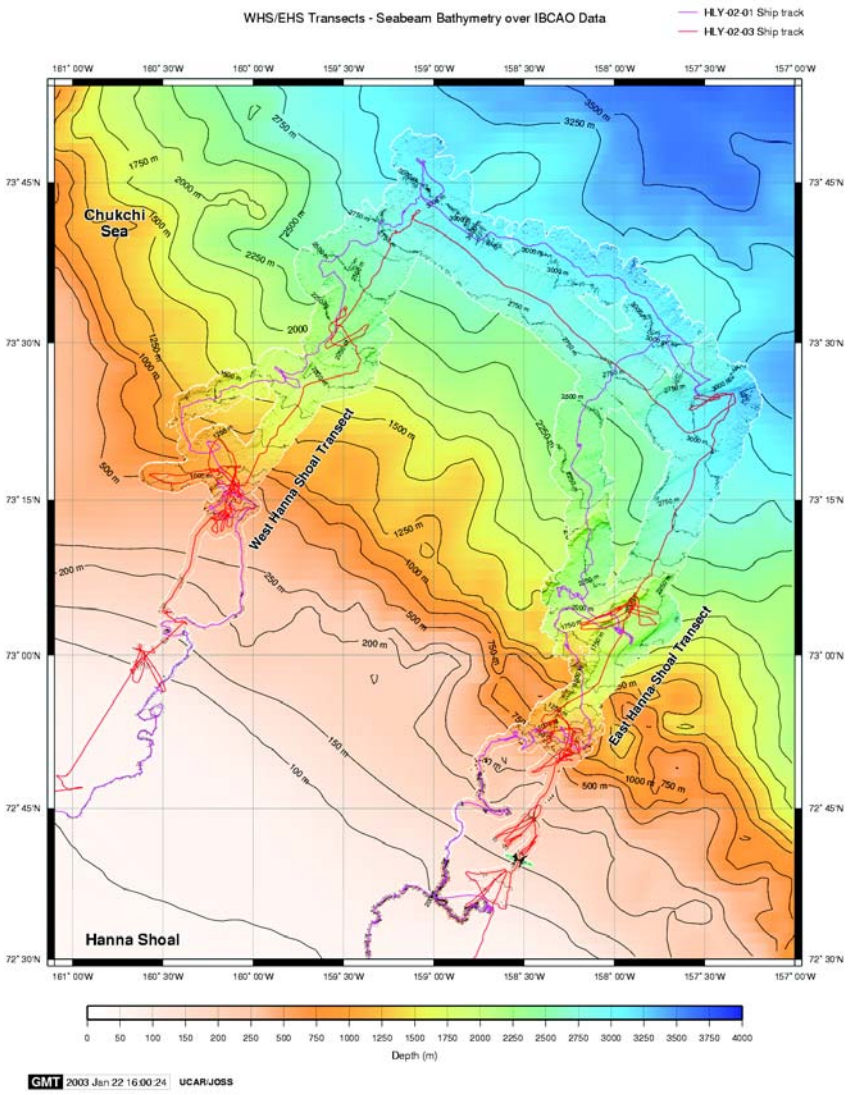
As part of JOSS's data management services during each cruise, several data sets were also collected and archived to tape for later availability through the SBI data archive at JOSS (http://www.joss.ucar.edu/sbi). Among these were satellite data pass files for each overpass of the ship by NOAA and DMSP polar orbiters. Information from the ship's log, including hourly weather conditions and operational times and locations, was also collected for the archive.

The complete field catalogs for both cruises are now online at JOSS (http://www.joss.ucar.edu/sbi/catalog). Consistent with the policy for NSF-sponsored research data residing at JOSS, research-quality data sets collected for SBI will be password protected for project participants for up to 2 years after collection to allow time for analysis and publication.

Data Management Web Sites for ARCSS Field Projects

JOSS works with project participants to establish data-management-oriented web sites that provide a single point of access to data and data-related products produced by the project. All JOSS-supported ARCSS field projects have web sites that can be accessed at http://www.joss.ucar.edu/arcss. Useful features common to all the sites include easy reference to the data management plan, data format and documentation guidelines,

The ATLAS Data Archive Page at JOSS. This page, located at www.joss.ucar.edu/atlas, is the gateway to ATLAS research data sets and supporting data at JOSS.



High-resolution Seabeam bathymetry data from the USCGC Healy *cruises (HLY-02-01 and HLY–02-03) overlaid on IBCAO bathymetry chart data in the same region. The two cruise tracks are shown by the red and magenta lines. Ship tracks that appear overlapping or confused are periods of free drift required for proper sampling. The thin black lines are isobaths. Note the fine-scale bottom details along the ship track derived from the Seabeam data.*

and simple data submission instructions. Access to the project data sets is available from the web page in several ways, including by year, by field site or cruise as appropriate, by discipline type, and by principal investigator. Other related links for access to distributed archive sites, project offices, or other relevant locations are provided as necessary.

Value-Added Data Processing

Bathymetric data from the Seabeam instrument were collected for each SBI cruise, along with data from the Knudsen and Bathy 2000 instruments. Maps were produced after decoding and replotting the Seabeam data, showing how much additional information about bottom characteristics can be available from these data. Currently available Arctic Ocean bottom topography from the International Bathymetric Chart of the Arctic Ocean Project provides the map background. This information will be crucial for the SBI scientists to interpret bottom samples taken during the SBI cruises.

JOSS can provide specialized support for the quality control of certain types of data collected for or by the field projects. JOSS worked with the ATLAS scientists and the NCAR Atmospheric Technology Division to perform checks on sounding data taken at the field site in Council, Alaska. JOSS has established specialized upper air sounding processing to help assure quality and uniformity in these data. JOSS received upper air sounding data from a variety of sources in multiple data formats and developed software to read and convert all of the soundings to the same columnar ASCII format. Three types of automated quality checks are applied to the sounding data. The checks confirm that the format conversion step was properly completed, verify that the data are within reason for the earth's atmosphere, and examine the vertical consistency of the sounding. None of the actual data within the sounding are changed, but quality-control flags are applied to specific data points and included with the sounding data. JOSS also visually examines a random portion of the sounding data to see if there are any consistent problems that the automated checks do not catch (such as humidity calculation problems). The data are then made available to the community via the Internet.

Consistent Project Data Format and Documentation Guidelines

One of the big challenges facing any field project is to collect the data in such a way that it permits simple and timely exchange with fellow participants and eventually with the larger science community. Considerable time is spent working with project scientists to reach agreement on data format and documentation guidelines. The following key components make up a successful data format structure:

- A consistent data file naming convention is defined so that files have unique identification, even for similar measurement types. Using an extension (such as .jpg or .txt) helps others users recognize the file format.
- Header record information includes contact information, temporal and spatial coverage, type of platform (such as ship or aircraft), coordinates (latitude and longitude), data ver-

sion number, and any other details of specific importance to that data set (such as measurement thresholds or a missing flag).
- Each row of data record includes a date and time stamp, the position, and the data. Alternatively each column following the date, time, and position stamp could be a parameter (with units) when multiple measurements are made at a single place or time.

The documentation that accompanies each project data set is as important as the data itself. This information permits collaborators and other analysts to become aware of the data and to understand any limitations or special characteristics of the data that may impact its use elsewhere. The data set documentation should accompany all data set submissions. While it will not be appropriate for each data set to have information in each documentation category, JOSS's guidelines should be adhered to as closely as possible to make the documentation consistent across all data sets. It is also recommended that a documentation file submission accompany each preliminary and final data set.

Project Data Documentation Guidelines

TITLE: This should match the data set name

AUTHOR(S):
-Name(s) of PI and all co-PIs
-Complete mailing address, telephone/fax numbers, web pages, and E-mail address of PI
-Similar contact information for data questions (if different than above)
-Grant number and title

DATA SET OVERVIEW:
-Introduction or abstract
-Time period covered by the data
-Physical location of the measurement or platform (latitude/longitude/elevation)
-Data source, if applicable
-Any World Wide Web address references (additional documentation such as the project's WWW site)

INSTRUMENT DESCRIPTION:
-Brief text (1–2 paragraphs) describing the instrument with references
-Figures (or links), if applicable
-Table of specifications (such as accuracy, precision, or frequency)

DATA COLLECTION and PROCESSING:
-Description of data collection
-Description of derived parameters and processing techniques used
-Description of quality control procedures
-Data intercomparisons, if applicable

DATA FORMAT:
-Data file structure, format, and file naming conventions (for example, column-delimited ASCII, NetCDF, GIF, JPEG)
-Data format and layout (description of header/data records, sample records)
-List of parameters with units, sampling intervals, frequency, range
-Description of flags, codes used in the data, and definitions (such as good, questionable, missing, estimated)
-Data version number and date

DATA REMARKS:
-PI's assessment of the data (disclaimers, instrument problems, quality issues, missing data periods)
-Software compatibility (list of existing software to view/manipulate the data)

REFERENCES:
-List of documents cited in this data set description

Preparation of Specialized Analysis and Outreach Products

As part of its support to ATLAS, JOSS worked with project scientists to compile the first of several data CDs that will be used to share useful data and information from the project with Arctic researchers, as well as educators, students, and other interested users. The initial CD is a compilation from the Ivotuk site on the North Slope of Alaska. The introductory web site contains data, photos, and descriptions gathered by more than 30 scientists and technicians encompassing a 2.5-year period from early 1998 through June 2000. The main purpose of the CD is to provide a single archive source for the multidisciplinary data collected at this site, in addition to presenting an overview of the project for those interested but not conversant with the individual disciplines.

The CD was designed to be operate using Windows or Macintosh and most web browsers. The principal features of the CD are:
- A self-contained archive on a single CD;
- Interactive site maps;
- Overview information on site and group activities;
- Cross-referencing of data by site, group, discipline, and year using menus, tables, and maps;
- Detailed data documentation; and
- A slide show and other sequences of interest.

Considerations for the Future

The Arctic research scientist of the future will be able to deploy or redirect assets that are mobile, long range, and easily relocated. They will want more data in near real time to assist with the monitoring and assessment needed to use deployed facilities and instrumentation more effectively and safely. The need and availability of data and

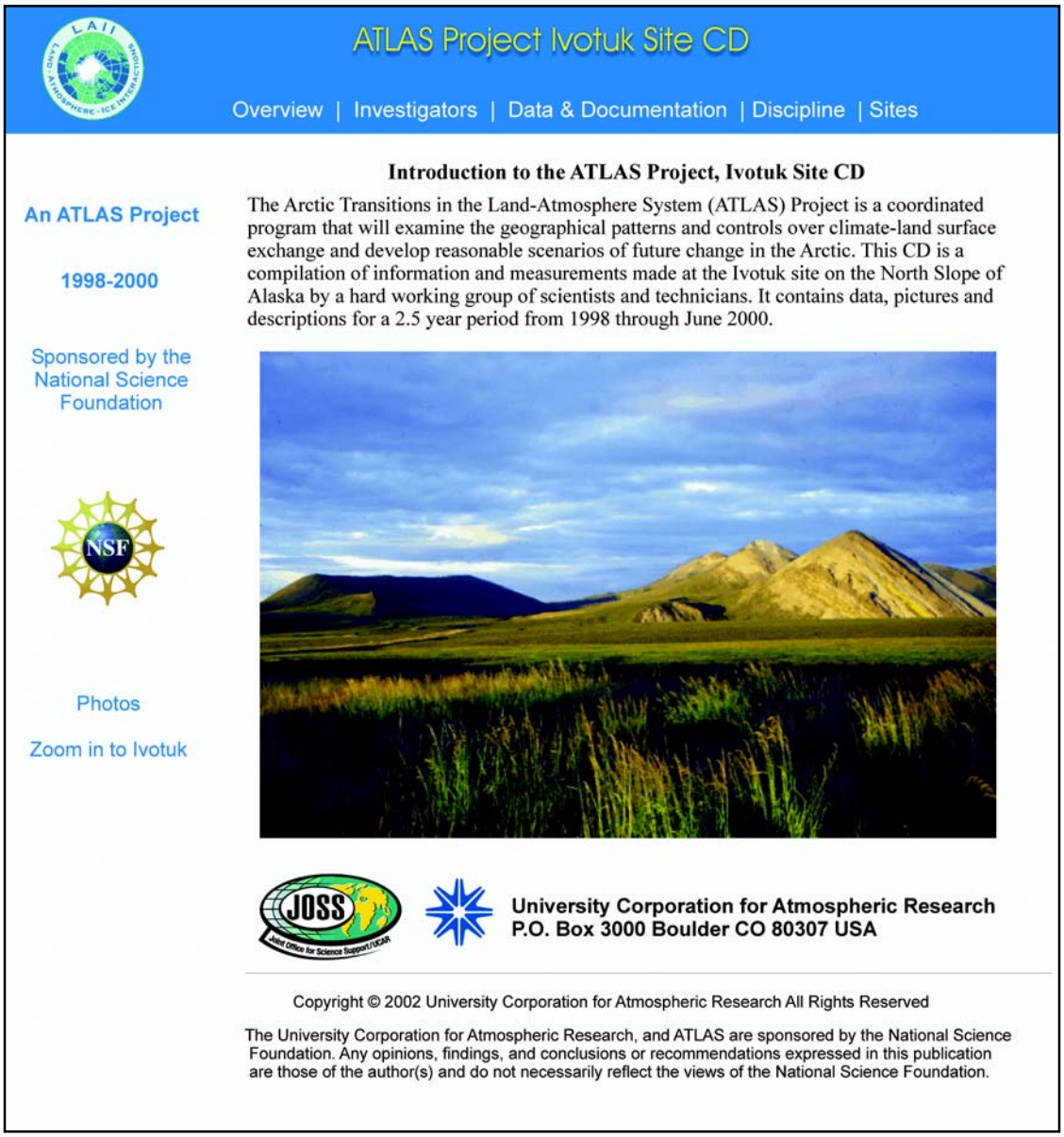
products in real time during field campaigns will increase as researchers conduct more complex experiments and deploy facilities and instrumentation remotely. Remote campaigns in the future can be linked directly to the classroom involving students at all levels in the scientific enterprise.

Data interoperability, or incorporation of data into a single analysis or visualization environment from distributed archives, will be standard for earth science researchers.

The researcher of the future will have seamless web access to project data, using high-bandwidth network connections and powerful data visualization, retrieval, and analysis tools. JOSS seeks to develop these tools and capabilities in its data management system.

There will be an increasing interest in multi-disciplinary synthesis of data sets as the research community scales up beyond basic research to address regional and global climate questions. The science community must be able to draw in data sets efficiently and reliably from distributed data archives to create the analysis data sets needed to address these important questions.

JOSS will continue to work hard to support the efforts of ARCSS and provide a phased approach by providing continuing assistance to new projects such as SEARCH and the Freshwater Initiative. They will take every opportunity to implement new technologies, matching the requirements and capabilities outlined above in an effort to aid the scientists in achieving their science objectives.



The ATLAS Project Ivotuk Site CD page at JOSS. This page, located at www.joss.ucar.edu/atlas/ivotuk_CD/html/IvotukFrameset.htm, is the online interface to the complete CD. Interactive maps, a zoom movie, annotated slide shows, and many data plots are some of the features of the CD.