Suggestion for the Data sub-group of the AOSB/MWG

Development for an Arctic data collection and sharing system (DCSS)

1. Background

The difficulty in data sharing arises mainly from two issues. First, each country imposes its own limitation on the release of data and second, the nature of the Arctic marine data.

Researchers who intend to submit data for sharing must address the following problems:

(1) Metadata. It is not always clear how much metadata is sufficient for data centers and researchers.
(2) Protocols. Research protocols vary between countries.
(3) Format. Researchers are not always familiar with the variety of data formats.

These issues serve as an impediment to the submission of data. Researchers often do not know how to submit data, how much time to spend on the submission, and how to address problems that arise from the data they have submitted. Many, therefore, give the data to data management experts who often do not understand the data and will never use the data.

2. Data collecting and sharing system (DCSS)

Faced with these issues, a possible solution is the design and development of software to solve the problems and difficulties outlined above. The software must serve four functions:

(1) Data collection. It should be possible to bring the software to the field (ship and other in situ conditions). After the data is obtained, the researchers could firstly upload the data into DCSS. The DCSS will shift the data into a common format which is accepted by all data management institutions. Meanwhile, the DCSS will ask for metadata. The researcher in the field will input all the required metadata. The DCSS will examine the data initially to ensure the quality of the data. If there is any problem, the DCSS will refuse to accept the data and ask the owner of the data to solve the existing problems. If the data is samples and need to be analyzed after the cruise, the DCSS will collect metadata first to achieve the data information and wait for the input when the sample is analyzed. With the aid of DCSS, researchers can easily collect and input data without the help of a data management expert. A summary of

the data will also be collected to record the contribution by data collectors. In this way, the basis for data sharing is built.

(2) Data analysis. The DCSS software could have a capacity to analyze and process various types of data. The function is somewhat like the Ocean Data View (ODV), but extends to more kinds of data. When the software has this capability, then researchers will wish to collect data by DCSS.

(3) Data sharing. With this function, people can easily use the data collected by other people with the same DCSS system. The software provides the same language to all scientists for same kind of data. It makes it easier to share data among scientists, between scientist and data centers, and data exchange between data centers.

(4) Data submission. Because the data has been stored and managed by the software, it is easier to submit data. When the owner of data gets permission from his or her authorizing agency, he could submit the data with the click of a button.

This DCSS system will need to be upgraded from time to time and researchers will use the system to manage and use data they have. Hopefully, with this DCSS system, the objective basis for data sharing will be built. If the software is built by the international effort, it will be an authoritative system to be accepted by most scientists.

3. Structure of the DCSS

The DCSS will be built with two parts: one is a**platform**, and the other is **modules**.

The platform is core software, which will be operational for users. It includes the functions to manage data bases, to connect modules, to pursue data analysis, and to process figures and images.

Because there are many kinds of data, the system must have different modules. The modules could be designed by data management experts, the researchers, or any people who are interested in the task. The modules could be easily assembled on the DCSS platform if they are designed according to the demanded format. The modules should include the data format, metadata content and format, and the database administration. The modules will be examined by a software center and issued through the website. The users could download the modules and assemble the data on their own DCSS platform.

4. Data types

(1) Ocean data

- Physical oceanography data: CTD, ADCP, optics, turbulent, turbidity, waves, etc.

- Chemical data: nutrients, metals, tracers, organics, POPs, etc.
- Biological data: bacteria, microalgae, ice algae, phytoplankton, zooplankton, benthos, fishes, mammals, birds
- Ecological data: oxygen, Ph, chlorophyll-a, PAR
- Geophysical data: gravitation, magnetic, earthquake, lateral scanned sonar, etc.
- Geological data: surface deposit, core, etc.
- Geochemistry data: CO2, DOC, DIC, POC, PIC, etc.
- Mapping data: acoustic sonar, multi-beam, imaging
- Air-sea coupling data: wind speed and directions, pressure, air temperature, humidity, gradient wind, cloud, radiation
- Ice-sea coupling data: ice coverage, ice concentration, ice thickness, melt ponds, albedo, leads, polynias
- Buoy data: drifting buoy, profiling buoy, mass balance buoy, mooring buoy, drifter, sub-drifter, up-looking sonar, etc.
- Mooring data: deep mooring, shallow mooring, trap mooring
- Remote sensing data: satellite images, helicopter digital photos, ship digital images, infrared images, X-band radar data.
- Retrieved data: all retrieved data from remote sensing images.
- Assimilated data: all assimilated parameters.

(2) Atmospheric data
(3) Snow and ice data
(4) Other data