Data Management Best Practices Ad-Hoc Committee
Report to the UNOLS Council, October 11, 2007

Stephen Miller and Robert Arko, Co-Chairs

An ad-hoc committee was established in July 2007 to report to the UNOLS Council on current community-wide practices in data and metadata capture when collecting data at sea, and to identify best practices and make recommendations for improvements. The membership has been selected by the UNOLS Chair to represent a wide spectrum of disciplines and ship operations across the UNOLS community. The members are:

Arko, Robert	Lamont-Doherty Earth Observatory, Columbia University
Case, James D.	University of New Hampshire
Chandler, Cyndy	Woods Hole Oceanographic Institution
Chapman, Piers	Texas A & M University
Fanning, Bill	University of Rhode Island
Forcucci, David	US Coast Guard
Graybeal, John	MBARI
Johnson, H. Paul	University of Washington
Measures, Chris	University of Hawaii
Mendelssohn, Roy	NOAA/NMFS/SFWSC
Miller, Stephen	Scripps Institution of Oceanography, UCSD
Smith, Shawn	Florida State University

During August and September 2007 the following results have been achieved by the Committee:

Organization. A web site has been established (http://data.unols.org) to allow member to exchange information. It currently contains Committee reports and work-in-progress documents, funding agency mandates, and related community initiatives. An email alias, data_manage@unols.org, is used to communicate with the group, along with the UNOLS Chair, Office, and Executive Secretary.

Charter. As requested by the UNOLS Chair, the Committee has drafted a Charter for review at the October 11, 2007 UNOLS Council Meeting. The Charter is posted on line (http://data.unols.org/charter) and contains the following basic charges to the Committee:
  a. Define the scope of platform and data activities.
  b. Address user concerns.
  c. Define UNOLS Community goals.
  d. Review available technical capabilities and procedures.
  e. Make recommendations.
  f. Maintain communications with the UNOLS Community.

Initial Survey. A survey was conducted among the members of the Committee in September 2007 to begin to understand the data needs and problems facing the UNOLS community, and to categorize the experience represented by the members. The results of the survey and an initial analysis are posted at http://data.unols.org/initial-survey-of-data-management-issues. Across the Committee, it appears that the highest priority is "shipboard metadata creation." The ranking indicates the severity of the problem and the need for a solution. It is to be noted that just

because an issue is not ranked as a top priority doesn't mean that it is not important. Rather, it may be seen as less of a technical challenge.

UNOLS Data Management Priority Issues

The top three problem issues are:
- **Shipboard metadata creation,**
- **Shipboard data quality control,**
- **Shore based quality control.**

The mid-level problem issues are
- **Dissemination to repositories,**
- **Shipboard archiving and copying,**
- **Shipboard access to data by scientists,**
- **Shore based access to data by scientists.**

Additional issues are
- **Data transport to shore,**
- **Dissemination for foreign clearance requirements,**
- **Dissemination to chief scientists.**

Meetings. The Project Kickoff teleconference was held on September 28, 2007 to allow members to introduce themselves and express what they perceive to be the most important issues. The Committee confirmed the scope of UNOLS activities to be addressed, and the communities of users that are represented. The Charter was discussed, revised and approved for submission to the Council. Relations to other activities were discussed, including the metadata survey by Shawn Smith, which will be presented at the RVTEC Annual Meeting November 6-8. The teleconference agenda and minutes are posted at http://data.unols.org/meetings/2007-09-28/.

The next Committee meeting will be held at the AGU Meeting in San Francisco, tentatively planned for noon on December 12, 2007. The agenda will include a discussion of proposed minimum standard UNOLS metadata and data products. An additional Committee meeting may be held in conjunction with the Ocean Sciences Meeting and the Marine Metadata Interoperability Workshop in Orlando, during the March 1-7 timeframe. Much of the work of the Committee will be conducted via email and the web site, although monthly teleconferences may be needed to make sure that the group is moving ahead with a consensus and a focus.

DATA MANAGEMENT BEST PRACTICES AD-HOC COMMITTEE

Terms of Reference
Revised September 28, 2007

1. INTRODUCTION

The UNOLS Council has requested the operation of an ad-hoc committee to report on best practices for the collection of data and metadata at sea to promote public dissemination. The primary issue is that federally-funded investigators are increasingly being required to make their data available within prescribed time periods to the community at large. That community will likely expand from other researchers to include managers and the public. UNOLS should play a significant role in facilitating that dissemination by ensuring that the collection of the data, samples, and metadata at the shipboard end of the process is done in such a way as to efficiently and effectively promote the information dissemination in compliance with whatever policies might be established by the sponsoring agencies or other authorities (e.g., permitting agencies). Furthermore, early establishment of uniform UNOLS standards and formats for data dissemination might avoid the proliferation of multiple disparate national standards and formats for meeting foreign clearance conditions, as well as serve as a template for ocean observatories.

At the March 2007 meeting of the UNOLS Council, the decision was made to establish an ad-hoc committee to report to the UNOLS Council on current community-wide practices in data and metadata capture when collecting data at sea, and to identify best practices and make recommendations for improvements. Existing data policies and dissemination mandates need to be identified, and an assessment made of how well the community is doing in meeting those requirements. This committee will be asked to refine their own charter in keeping with the issue as outlined above, and report back to Council with a brief White Paper on the subject within one year's time. If necessary, the Council may extend the activities of the Committee beyond one year. Volunteers to serve on the committee were solicited from the community and agencies.

2. MEMBERSHIP
The membership has been selected by the UNOLS Chair to represent a wide spectrum of disciplines and ship operations across the UNOLS community. The Co-Chairs are Stephen P. Miller (SIO) and Robert Arko (LDEO). Changes to membership will be made by the UNOLS Chair, as he/she deems appropriate.

3. MEETINGS
It is expected that much of the work of the Data Management Committee will be conducted via email and teleconferences. Discussions may be held at national meetings for those members who are attending. The Committee and the Chair of the Council will evaluate the need for a face-to-face meeting as the effort progresses. An email alias, data_manage@unols.org, is used to communicate with the group, along with the UNOLS Chair, Office, and Executive Secretary. Documents will be posted for sharing on a content-managed web site http://data.unols.org.

4. CHARGE TO THE COMMITTEE

After the Committee begins its work it may refine a series of goals and milestones, and report back to the Council. Initially, the goals of the Committee are:

a. **Define the scope of activities.** The Committee will address those elements of ocean information that are now or will likely be under UNOLS influence either by virtue of being a UNOLS platform or being equipment deployed from a UNOLS platform (AUVs, ROVs, towed vehicles, etc.) The Committee will need to address routine underway observations and samples, as well as procedures for managing metadata and data from special equipment brought on board by scientific parties. The activities will also include long-term access to data derived from UNOLS systems, interactions with repositories, and responses to data requests from foreign governments. Future activities may be identified. At the present time, data from satellites and observatories are not included, as they are addressed by other organizations. The "UNOLS Community" is understood to include UNOLS vessel and platform operators, data generators (original producers, and also creators of derived products), data managers (national, institutional and project repositories), and scientific researcher and student end-users.

b. **Address user concerns.** The Committee will conduct a "top-down" review of the technical capabilities, procedures, and challenges of the UNOLS data systems. The relevant communities of interest shall be listed and described, covering the appropriate scientific disciplines, data activities, and platform operations, and representatives identified to speak for each aspect. An initial survey shall be conducted to assess data needs and problems facing the UNOLS Community. Additional responses from users beyond the Committee members may be solicited as necessary. Common themes across the community shall be identified, as well as important diverse needs. Any new requirements (e.g., for additional data reporting mandates) shall be identified and documented. The existing situation will be characterized to provide context for recommendations. The Committee shall summarize the results, and assign overall priorities to issues.

c. **Define UNOLS Community goals.** The Committee shall identify a series of data "best practices" goals for data management to guide the choice of potential solutions to problems.

d. **Review technical capabilities and procedures.** The Committee shall compile a list of possible solutions to the problems and make recommendations, based on an understanding of the technology and an awareness of other activities in development, with an emphasis on the highest-level priorities. It is anticipated that this will stimulate a discussion of cruise metadata reporting, interactive web form practices, management of controlled vocabularies, and repositories.

e. **Make recommendations.** Based on its review of UNOLS activities, user concerns, and available technologies, the Committee will make a series of recommendations. Broad "top-down" guidance documents will be developed, with suggestions for "best-practice" approaches for every type of operator. Capability enhancements may address software, hardware, automated services or resource improvements to improve data management. Procedural improvements may address changes in human or automated processes that improve data management outcomes. Policy recommendations may highlight ways to improve UNOLS data access and timely dissemination.

f. **Communications.** The activities and decision making processes of the Committee shall be transparent. Input may be solicited from an audience broader than the Committee membership. Reports and other information may be posted on a public web site and broadly disseminated by other means, as appropriate. Reports shall be made to the UNOLS Council at regularly scheduled Council meetings. The Final report will be provided to the Council prior to their late spring/summer 2008 Council Meeting.

Adopted: October 11, 2007, Arlington, VA