# WOODS HOLE OCEANOGRAPHIC INSTITUTION

---

# FINAL REPORT OF THE AD HOC SCIENTIFIC DATA ADVISORY COMMITTEE

## 12 MAY 1999

**COMMITTEE MEMBERS:**

Brian Tucholke, Chairman (Department of Geology and Geophysics)
Richard Chandler (Marine Operations)
Ruth Curry (Department of Physical Oceanography)
David Glover (Department of Marine Chemistry and Geochemistry)
Robert Groman (Department of Biology)
Jonathan Howland (Department of Applied Ocean Physics and Engineering)
Colleen Hurter (MBL/WHOI Library)
Andrew Maffei (Computer and Information Services)

**EX OFFICIO:**

Cathy Norton (Director, MB./WHOI Library)

**MBL/WHOI LIBRARY COMMITTEE LIAISONS:**

Daniel Fornari (Department of Geology and Geophysics)
Mark Kurz (Department of Marine Chemistry and Geochemistry)

**NSF OCEAN DATA POLICY AT:**

**http://www.geo.nsf.gov/oce/**

## EXECUTIVE SUMMARY

The ad hoc SDAC arrived at the following conclusions and recommendations.

- WHOI has a fundamental responsibility to collect, archive, manage, and distribute important scientific data.
- WHOI should institute a proactive policy for archiving scientific data that are acquired by WHOI scientists and by WHOI ships and deep submergence vehicles.
- To the extent reasonable and possible, scientific data should be archived with national data centers and other established archives. Other scientific data, including those data desirable for local access, should be archived at WHOI, following a priority list based on the scientific value of the data. NDSF data are at the head of this list.
- Implementation of this policy will require personnel, internal adjustments, and physical resources as follows:
  - A Scientific Archivist to track, acquire, and help manage scientific data for both the Data Library and the Seafloor Samples Laboratory.
  - A Cataloger, a Clerical Assistant, and a Mixed-Media Preservationist to manage and rescue existing data, and to manage future acquisitions within the Institution Archive.
  - Adjustment of procedures in Marine Operations so as to require both complete Cruise Prospectuses and Cruise Reports from Chief Scientists in a timely manner.
  - Adjustment of WHOI SSSG technician duties on each WHOI cruise to assist in data handling and documentation that will facilitate archiving functions.
  - Provision of standard "tools" (e.g., hard-copy logbooks, possibly software) by the Institution to scientists in order to document scientific data and to assist in archiving.
  - Active, annual promotion by the Institution of the value of archiving protocols and of archiving scientific data by the WHOI staff.
  - Establishment of a permanent Scientific Data Advisory Committee to deal with continually evolving issues of acquisition, archiving protocols, management, and dissemination of scientific data. Establishment of mechanisms routinely to cull unwanted data from the archives and to identify data that need to be rescued or migrated. Provision of support to staff who contribute to these functions.
  - Construction and use of an efficient, WWW-accessible metadata (data about data) system (hardware and software) to manage scientific data and make it accessible to the scientific community. Incorporation of appropriate web links to other archives and digital databases.
  - Creation of new space to meet the needs of both the Data Library (at least 2600 square feet over the next 20 years) and the Seafloor Samples Laboratory (approximately 2400 square feet over the next 20 years). Each of these facilities should remain centralized in one location.
  - Upgrade of DSOG hardware and procedures to allow complete duplication and archiving of NDSF data.
  - Development of a mechanism similar to the Independent Study and Sr. Technical Staff Awards to award internal, proposal-based grants dedicated to enhancing the Scientific Archives.
  - Investigation of mass-store systems for large volumes of high-priority digital data.

## I. INTRODUCTION

The ad hoc SDAC was charged with providing advice and recommendations on long-term (5-20 year) strategies and infrastructure requirements for retention, storage, migration, and retrieval of scientific data to meet immediate and future needs of the Institution in carrying out its mission and its responsibilities to the science and engineering communities. A non-exclusive list of topics to be considered included a) existing policies for scientific data archiving, b) legal requirements for archiving associated with grants and contracts, c) requirements associated with national facilities (e.g., Deep Submergence, NOSAMS), d) policies of scientific journals associated with archiving of data, and e) services required/desired by WHOI's scientists and engineers.

The ad hoc SDAC met seven times between September 1997 and January 1998 to consider the archiving needs of the Institution (Appendix I). During this period members of the ad hoc SDAC also convened department meetings, met with individual scientific and technical staff, or otherwise surveyed department members (e.g., by questionnaire) to obtain input on procedures and policy for archiving scientific data. A summary of pertinent comments from these sources is included in Appendix II.

The meanings of acronyms used in this report are listed in Appendix IX.

## II. INITIAL CONSIDERATIONS

The ad hoc SDAC devised a set of working definitions, agreed on a scope of data to be considered, and adopted a set of guiding principles.. From the outset, we recognized that dealing fully with the tremendous range of possible archiving issues is beyond the capability of a single, short-lived ad hoc committee. The following defines our perspectives and the general limits of our analysis.

### A. Archives - Definitions

The central facility around which most of our discussions revolved is the "Data Library" in the McLean basement. This archive contains a large volume of scientific data, including deep-submergence data ("Alvin Archive"). It is part of the "Institution Archive" which also includes other Institution data such as administration and corporation papers. We restricted our deliberations almost entirely to scientific data, including scientists' papers that have been given to the archive. The "Seafloor Samples Laboratory" located on the main floor of the McLean building is also an archive of geological samples (cores, dredges, etc.). We considered both the Data Library and the Seafloor Samples Laboratory in our deliberations and collectively refer to them as the "Scientific Archives". The present status of these archives is summarized in Appendix III.

Beyond these archives, there are a host of facilities that we lump together under the term "distributed archives". They fall into two classes. "WHOI distributed archives" consist of various data repositories presently maintained by WHOI groups or individuals and that are, or may become, accessible on the World Wide Web; some of these may have a limited life at WHOI. Examples include U.S. GLOBEC Georges Bank, JGOFS, and NOSAMS (see Appendix IV). We have not made an exhaustive survey of these repositories; we discuss them generically with occasional specific examples. "Global

distributed archives" are data repositories, largely web-accessible, outside the Institution. They include national data centers (NODC, NGDC, etc.), university repositories (e.g., the LDEO database), and other similar repositories.

**B. Scope of Scientific Data Considered by the ad hoc SDAC**

In general, scientific data considered by the ad hoc SDAC are limited to data acquired by WHOI investigators, and to data acquired by non-WHOI investigators while using WHOI facilities. The data acquired by WHOI investigators include data obtained at sea (WHOI and non-WHOI vessels), in the field (on-land and coastal environments), and in the laboratory.

**C. Categories of Scientific Data**

We consider scientific data to fall into two basic categories. Primary scientific data are data that, if lost, 1) cannot be replicated (examples: raw or original data, data on time-variable parameters), or 2) can be replicated only at great expense (examples: processed data, certain kinds of sample analyses). Secondary scientific data are data that can be replicated with limited effort or cost (examples: syntheses of data; data tables that are published, are in technical reports, or are otherwise available).

As a guiding principle, the ad hoc SDAC considers that the more difficult it is to replicate a given kind of scientific data, the more important it is to archive that data.

**D. Who are the Scientific Archives for?**

Before developing guidelines for archival policy and archiving mechanisms, it is necessary to understand who the Scientific Archives serve. The best answer to this question comes from surveying the historical use of the Data Library and Seafloor Samples Laboratory. However, we recognize that future changes in archival policy and archiving mechanisms may well affect both the balance of user groups and the degree to which the Scientific Archives are used.

The primary users of the Data Library are WHOI scientists and WHOI/MIT Joint Program students. The Data Library also serves the other members of the Woods Hole scientific community, e.g., USGS, SEA, BUMP students, MBL, and NMFS.

A second group of users consists of researchers from outside the Woods Hole area including universities and government agencies, and especially people from private companies which comprise a large percentage of this group. Many of these patrons are former WHOI staff and students who retrieve data that they may have collected while they were here. Other outside patrons are those who have collaborated with WHOI scientists.

A third group of users has developed from "word-of-mouth" knowledge of our collection. The collection of reports, atlases and maps is slowly becoming known because it is now partially cataloged, and information about the holdings is on the library web page. The international ocean science community uses the facility for history of oceanography, accessing the Alvin Archive, and other general scientific endeavors.

The general public is not yet a consistent user of the Data Library, but as on-line information about the collection grows, the scope of the service definitely will change. It is expected that this group will be a force that will impact service in the areas of education and outreach on the part of the archive staff.

Users of the Seafloor Samples Laboratory historically have been roughly equally distributed between WHOI scientists and outside investigators, although WHOI scientists have become the majority users in recent years. There is significant use of samples by investigators at nearby institutions (e.g., USGS, SEA, MIT, BU), but larger use by investigators across the U.S. and internationally. In 1994-96, samples distributed to international users accounted for nearly a quarter of the 8400+ samples distributed outside the Institution. WHOI scientists during this period received more than 12,000 samples.

**E. Responsibilities of WHOI**

WHOI, as a world leader in oceanographic research and a significant consumer of public research funds in the ocean sciences, has a fundamental responsibility to acquire, preserve, and disseminate scientific data. Much of the function of dissemination is accomplished by normal publication by the scientific and technical staff. Primary scientific data that are not otherwise generally available to the community should be sought out, preserved, and disseminated by the Scientific Archives. This function is becoming increasingly important because funding for field programs is tightening, researchers are looking to historical data as a practical matter (i.e., because of limited funding) and for new insights (e.g., into climate change), and funding agencies are becoming more critical about the responsibility of the research community to make their scientific data widely available.

The ad hoc SDAC feels that WHOI must take its archiving responsibilities seriously. If the Institution acknowledges and accepts this responsibility, it must make a clear and continuing commitment to proper scientific archiving. This commitment consists of a) promulgating an archiving policy for scientific data that is consistent with requirements of funding agencies, fair to scientific investigators, and *proactive* in seeking to preserve important scientific data (see Section IV), and b) providing the resources to effectuate this policy (see Projected Needs in Sections VI and VII). Consistent and comprehensive data archiving policy and practice will improve awareness of and access to WHOI scientific data for researchers both at the Institution and in the larger scientific community. By supporting the archives and making firm, long-range commitments, WHOI will also improve the attitude of the "local culture" about the importance of archiving scientific data.

**F. Rationale for Archiving Scientific Data**

Considering the values of scientific data and the WHOI responsibilities noted above, the ad hoc SDAC considers that archiving of primary scientific data is important for the following reasons, roughly in order of descending significance:

- To assure that the data are properly documented and permanently accessible to the scientific and engineering communities.
- To make data available for purposes of education and public outreach.
- To meet requirements of funding agencies.
- To obtain funds from commercial rights and to apply these funds fully to maintaining and improving the Scientific Archives.

**G. Impact of the Technology Revolution**

There has been a dramatic change over the past decade in our ability to digitally acquire, process, store and disseminate scientific data, and this revolution continues at a rapid pace, e.g., through the World Wide Web. This brings obvious advantages, as well as some outstanding disadvantages. A fundamental disadvantage is that capabilities of hardware, software, and recording media are changing and expanding so rapidly that it is difficult for any archive to adapt fully in terms of either manpower or physical resources. As a result, archived digital data may become unreadable in a short period of time. This problem cannot be solved, but suggested practices to help alleviate the situation are outlined in Section V.B.1.

Another major disadvantage is that it is nearly impossible to predict future technology changes, and thus to design the details of archival mechanisms and practices intelligently. Thus we have limited our recommendations to general principles, as outlined in Section V. We also recommend the establishment of a permanent SDAC (Section V.B.1); one of its functions would be to monitor evolving technology and help to guide archiving practices accordingly.

**H. Limitations of Resources**

The ad hoc SDAC was instructed to advise on archiving strategies and infrastructure without considering costs. Nevertheless, WHOI will have limited resources for archiving under the best of conditions, and it is prudent to take advantage of all opportunities to minimize costs while assuring that important data are archived. Thus we adopted the following philosophy for general methodology in archiving scientific data.

- For primary data:

1- Archive at appropriate national data centers whenever possible, eliminating the data from the Scientific Archives.
2- Archive at archives of other institutions when appropriate, eliminating the data from the Scientific Archives.
3- Utilize WWW-accessible archives whenever possible; WHOI should maintain metadata (data about data) pages to provide tracking and access.
4- Utilize the WHOI Scientific Archives for the remaining data; establish and encourage use of standards for digital data.

- For secondary data:

1- Archive as above when justified by value of data, convenience of access, or other factors.

Countering the emphasis of the above prioritized list, we recognize that there are certain kinds of data and metadata which are appropriate for WHOI to archive irrespective of whether they are also available elsewhere, e.g. in national data centers, or could be archived elsewhere. For example, the values of easy internal access, facilitation of data management, or simply Institution visibility/pride may outweigh the costs involved. Such data and access methods should be retained in the Scientific Archives so long as they can be reasonably supported.

The ad hoc SDAC notes that there are tremendous varieties and volumes of primary data generated by WHOI investigators and WHOI facilities (see Appendix V), so it may not be physically or financially possible to archive all these data properly. Consequently, the Committee developed priority-ranked lists of data for each department in Appendix

V, defining the general order in which data should be archived, managed, and (if need be) migrated or rescued.

**I. The WHOI Culture**

WHOI is a marvelous collection of entrepreneurs whose individual and collective activities are the mainstay of the Institution's vitality. Getting them to accept and promote consistent archiving practices will be like trying to herd stray cats. Nonetheless, the ad hoc SDAC feels that this culture can be beneficially modified over the long term by two actions: establishment of a consistent archiving policy, and promoting the policy and demonstrating that archiving is in an investigator's best interest. The former is outlined in Section IV, and the latter is discussed at the end of Section V.A.

## III. EXISTING SCIENTIFIC DATA POLICIES

### A. WHOI Policy

The ad hoc SDAC reviewed existing WHOI policy on archiving scientific data, which is included within Institution Memoranda #3-83 and #4-83 (Appendix VI). We identified three features of this policy that we consider to be problem areas. First, the tenor of the policy's philosophy is that data are to be *excluded* from the archive insofar as possible, in that the memoranda concentrate on how data are to be reviewed, excluded, or removed from the archive. The ad hoc SDAC feels strongly that the archiving policy for scientific data should reflect an attitude 1) that the archive wants to acquire primary data, and 2) that the archive will be proactive in acquiring such data. History shows that, left to their own devices, investigators at the Institution have a poor record of archiving primary scientific data with appropriate documentation, and a proactive archiving policy is necessary to counter this trend. At the same time, the archive must be selective about data that it will search for, accept, and retain; our proposed priority order for archiving scientific data is given in Appendix V.

A second problem with existing WHOI policy is that it calls for a "Data Library Review Committee" which was to make decisions on material that should be retained in the archive. As far as we can determine, this committee was never formed, so the archive has effectively been "adrift" in terms of receiving scientific and technical advice from the WHOI staff. We feel that it is essential to provide such advice in a routine manner if the Scientific Archives are to be effective and efficient in seeking, acquiring, managing, culling and disseminating scientific data. In Section V we propose mechanisms to accomplish this.

The third problem area lies with the specification in Memorandum #4-83, para. 2, of data to be archived. The list is incomplete (for example, there is a glaring omission of Alvin video and film data). Furthermore, at least some of these data are not routinely archived (e.g., Chief Scientist log) because there is no consistent mechanism for doing so. The ad hoc SDAC recommends that the archiving policy be more general on this topic. It also recommends mechanisms to assure that primary data are properly logged and archived (Sections IV and V).

### B. Policies of Funding Agencies

The primary agency guidepost for disposition of scientific data is that of the National Science Foundation (Appendix VI). NSF policy is that data are to be submitted within certain time periods to specified national data centers (e.g., NGDC, NODC, etc.) so that they become publicly available. The time period within which data are to be submitted is normally two years from acquisition, but for some kinds of data it is significantly less. It is the position of the ad hoc SDAC that any WHOI policy on archiving of scientific data and proprietary periods should be consistent with the policy of the agency that funded the research. This is reflected in our draft policy given in the next section.

We note here that the responsibility of providing scientific data to national data centers is assigned by agencies to Principal Investigators, not to the PI's institution or the institution that operates the vessels on which data were acquired. We feel that this is proper because it avoids making the Institution responsible for distributing data over which it may have no control.

We also note that while agency policies on disposition of scientific data have long existed, they generally have not been enforced. This is changing. With research funding becoming tighter, program officers are becoming more critical of investigators who neglect to make their data available to the full scientific community within the prescribed periods, and agencies are beginning to look toward enforcing their policies. This is actually beneficial to us in that it will help prod investigators to archive their data properly, both in the Scientific Archives and elsewhere. It also means that our internal efforts to review archiving policy and mechanisms are very timely.

## IV. PROPOSED WHOI POLICY ON ARCHIVING SCIENTIFIC DATA

The following paragraphs represent the basic philosophy of the ad hoc SDAC about archiving scientific data at the Institution. Mechanisms for carrying out this policy are outlined in the following section.

### A. General Policy Statement

Primary scientific data (or acceptable copies thereof) acquired by WHOI facilities or by WHOI investigators are to be archived in the Scientific Archives for the period of their useful life. In the alternative, such data can be made available to the scientific community via another acceptable, user-accessible facility (e.g., a national data center, a WWW URN [Universal Resource Name, a *permanent* WWW address], or archives at another institution). In this case, the Scientific Archives should be informed as to where the data are deposited and how they may be accessed, in order for the archives to maintain usable metadata files. Data are to be deposited with the Scientific Archives within 2 years from the time they were acquired (or less, if so required by the agency that funds the research) or as otherwise agreed between the archives and the Principal Investigator (PI). Data are proprietary to the PI during the period prior to required deposit in the archives.

WHOI also encourages investigators to deposit secondary scientific data with the Scientific Archives if these data have archival value and are not available to the scientific community by other means. If such data are deposited at another archive, the Scientific Archives requests that information be provided to them on methods of access.

In all cases, the Principal Investigator is responsible for meeting funding-agency requirements to archive appropriate data at a national data center. Staff at the Scientific

Archives are available to assist the PI in this process. If data are acquired in waters of foreign countries, the Principal Investigator is also responsible for providing data to the foreign countries as those countries may require.

Guidelines for archiving and retaining scientific data from various sources:
1. *Data collected by WHOI facilities* (e.g., ships and their data collection systems, including NDSF facilities). WHOI shipboard technical staff and the WHOI Scientific Archivist should have the responsibility to see that data are logged, WHOI-labeled, and archived at WHOI. When possible, data can be duplicated as desired by the PI, who will retain the copy. Data which are desired by the PI, but which cannot readily be duplicated, can be signed out to the PI, who commits to archive the data at WHOI within a time period mutually agreed on by the PI and the Scientific Archives; this period will normally be 2 years or as otherwise specified by extant policy of the funding agency.

2. *Data collected by WHOI PIs.* Data should be logged, WHOI-labeled, and signed out to the PI unless the PI wishes immediately to deposit the data in the Scientific Archives. After the time period as noted in item 1 above, data are to be archived at WHOI, or the Scientific Archives is to be informed as to how the data have been made accessible via a national data center, WWW URN, or other institutional archive. If the data were collected using non-WHOI instrumentation and/or non-WHOI facilities, the data should be logged and WHOI-labeled if appropriate. The data should be archived at WHOI after the time period noted above if they are not archived at another institution, national data center, or WWW URN.

3. *Data collected by non-WHOI PIs with their own instrumentation when using WHOI facilities (e.g., WHOI ships and vehicles).* Data should be logged as being collected, with notation that the data are in the possession of the PI. The PI is requested to archive the data at WHOI when analysis is complete, or to inform the Scientific Archives as to where data are archived.

4. *Data retention.* Archived scientific data in the Scientific Archives will be retained for the period of their useful life, managed, and (after the proprietary period) be made publicly available. Scientific holdings of the Scientific Archives will be periodically reviewed in order to identify and discard data that are no longer useful. Such reviews will involve appropriate representatives of the scientific and technical staff, the principal investigator who acquired the data (when available), and the staff of the Scientific Archives.

**B. NDSF Archiving Policy**

In late November, during the course of its deliberations, the ad hoc SDAC was asked for input on archiving policy for the National Deep Submergence Facility. Our proposed NDSF archiving policy is given in Appendix VII.

## V. IMPLEMENTATION

Once the ad hoc SDAC agreed on basic philosophy, priorities, and policy for archiving scientific data, we identified impediments to implementation of the policy and consolidated them into three problem areas. Each of these problem areas, with proposed solutions, is discussed below.

**A. Becoming Aware of Scientific Data Collection, Tracking the Data, and Acquiring it for the Scientific Archives**

To conduct its business of archiving, managing and disseminating scientific data effectively, the Scientific Archives must have mechanisms whereby it can learn what data will be collected, then track the data and eventually incorporate it into the archives. Currently, there is no reliable mechanism to accomplish this (notable exceptions are Alvin/NDSF and ship-navigation data). Consequently, the Scientific Archives do not consistently document and acquire primary data of high scientific and/or historical value. Such data typically are collected at great expense, but they commonly become fragmented, useless, or lost because of inadequate tracking and documentation. We recommend the following steps to deal with this issue.

- *Adopt and apply an effective Data Archiving Policy.* Our recommendation for the philosophy of a general archiving policy was given in Section IV.

- *Create a new position of "Scientific Archivist" within the Data Library.* This individual would have overall responsibility for working with scientists to track and acquire scientific data within the scope defined earlier. This would include working with personnel of the Seafloor Samples Laboratory to assure that all seafloor samples, whether or not archived at WHOI, are fully documented by metadata. The Scientific Archivist also would work with various scientific groups within WHOI to coordinate archiving efforts and establish linkages, and would work closely with the Data Librarian in managing and disseminating the data in the Data Library. This individual would be involved primarily with science-related aspects of archiving, and would report to the Senior Associate Director and Director for Research. Details of this position are given in Section VII.A.

- *Adjust the mechanics of certain WHOI internal operations to facilitate communication and to facilitate archiving.* This primarily will involve two groups: Marine Operations, and individual departments.

-Marine Operations (see also Appendix VIII). For the Data Library to become aware of and track the acquisition of scientific data on WHOI ships, it must have access to basic information on cruise plans, i.e., designation of the Chief Scientist, list of Principal Investigators, cruise track, and proposed data acquisition and distribution. This information should be gathered as a matter of routine by Marine Operations in the form of a Cruise Prospectus. Because the Chief Scientist often is lax about providing a Cruise Prospectus, Marine Operations must keep on top of the scientist to assure that this information is provided several months before a cruise. A Cruise Prospectus form is available on-line at http://www.marine.whoi.edu/webpub/cruplan/question.htm. The Scientific Archivist will access this information and contact PIs to work with them on the process of data documentation and archiving.

On each leg of a WHOI cruise, a SSSG or NDSF technician should be designated as the "Science Officer" and should perform two functions to assist with data documentation and archiving. First, he would provide standard "tools" to the scientific party (see below) to facilitate documentation. Second, he would work with the scientific party to document data during the cruise and to complete a Data Log by the end of the cruise. The Data Log would full document the data (e.g., what, where, when, etc.) so that the data will always

be useful to any interested party, and it would also record disposition of the data (i.e., signed out to the PI, or transmitted to the Scientific Archives) so that he data could be tracked. A well documented Scientific Log Book (see tools, below) may suffice to document data types, acquisition times, and so forth, but it must be accompanied by a list showing disposition of the data. The Science Officer should assure that originals of the following kinds of data are transmitted to the Scientific Archives: Ship navigation, data from ship systems (e.g., IMETS, SeaBeam, NDSF data, etc.), deck logs, Alvin and ROV logs, Scientific Log Book, and Data Log. The Chief Scientist is free to take copies of any of these data that he/she requires unless the data are restricted (e.g., SeaBeam data, Appendix V).

Marine Operations should also require a complete Cruise Report from the Chief Scientist at the end of a cruise. The Cruise Report should include overall scientific objectives and the purpose of cruise; scientific and technical personnel; summary of data acquisition systems and their performance; summary of data acquired and processing conducted at sea; identification of any problems with operations or facilities; ship track; and other information that the Chief Scientist may consider pertinent to documenting the cruise and the collected data for posterity (e.g., cruise narrative, winds and seas, operations teams, etc.). The one-page UNOLS summary "cruise reports" currently being collected and archived are *not* adequate for this purpose because they contain almost no useful information for documenting data. With the word processing and computational facilities now available on our ships, it should be an easy matter for a Chief Scientist to provide a complete Cruise Report in a timely manner. Marine Operations should send Cruise Reports to the Data Library for archiving with the cruise data.

-Individual Departments. It will be much more difficult to track, document, and acquire data for the Scientific Archives when the data are collected by WHOI scientists on non-WHOI ships or in field programs on land. A partial solution is to emphasize to the Scientific and Technical Staff that they should work with the Scientific Archives before they embark on such programs to assure that collected data are both well documented and subsequently logged by the archives. Another solution is for individual departments to keep track of upcoming field programs and to communicate this schedule to the Scientific Archivist. The Scientific Archivist will be proactive in seeking this information and will contact the scientist to provide "tools" and arrange logging and archiving of data. In each case, the WHOI scientist should be responsible for providing well documented records and data to the Scientific Archives.

• *The Scientific Archivist should search out, document, archive, and make available existing, unarchived scientific data within the Institution.* Practically speaking, this would have to be a long-term effort. It would also have a lower priority than implementation of the basic system to deal with tracking and acquiring newly collected data. However, the Institution should support the Scientific Archivist's efforts to interview investigators tenaciously and to document and archive data that investigators previously have collected or produced.

• *Provide "tools" to facilitate proper documentation and tracking of scientific data.* Standard "tools" will provide a mechanism to properly document data that may be acquired by the Scientific Archives. Foremost among these should be a Science Log Book (day, time, event, comments) that would be provided to all Chief Scientists on

WHOI ships, and to all WHOI scientists who are conducting field programs other than on WHOI ships. This log should provide "carbon" copies so that the copy can be retained by the scientist and the original can go directly to the archive. The advantage of such a log is that it is easy for the Institution to provide, it will always be readable, and it will rarely need to be migrated. WHOI used to provide these log books, but the practice has lapsed.

Another possible tool that might be provided is generic software that runs on PDAs, laptop computers, and workstations to catalog data and events, and to verbally document the conduct of the scientific program. This could be very useful on WHOI ships, notably when it comes to producing a cruise report at the end of the cruise. It might be less useful in other field programs (e.g., land-based programs).

The Scientific Archivist would be responsible for providing blank, hard-copy logs to WHOI ships and to WHOI scientists for other field programs. The shipboard Science Officer would make these logs and software tools available to the shipboard scientific party on WHOI ships.

- *Promote to the scientific and technical staff the value of scientific archiving.* The Institution should annually advertise and promote its policy on data archiving, just as it advertises policies on harassment, conflict of interest, and other issues. This annual message should also summarize the main holdings of the Scientific Archives and point out the archive facilities and tools for accessing and managing data, and it should encourage staff to use the archives. In addition, it should include information on the benefits that accrue to the scientific and technical staff when they archive their data. Among these benefits are: a) gaining the appreciation of funding agencies and program officers for properly meeting grant and contract obligations on data dissemination (this may translate into more favorable perspectives on proposals by program officers ), b) obtaining the good will of national data centers, which are much more amenable to providing data to investigators when they also receive data from them, c) opening new lines of communication with other scientists who access the data, which may lead to new collaborations or insights, and d) the possibility of financial gain if the data are commercially used (see Section V.C.4).

**B. Managing Data Within the Scientific Archives**

In order for data in the Scientific Archives to be accessible and usable, they must be managed effectively. To begin, the archives need to verify basic facts such as data type, when and where the data were acquired, and who's providing the data, as noted above. Effective management involves building procedures that provide for qualification, acceptance, possible reformatting, indexing, storage, migration to various media, and culling of unwanted data. The following identifies notable problem areas and suggests possible solutions.

**1. Data diversity**

There is a tremendous diversity of data types and recording media collected by WHOI researchers. These include paper records, samples (e.g., biological, geological), film, videotapes, and a wide variety of electronic media. It is difficult for any archive to maintain the resources necessary to assure that these data will always be readable. Such resources would have to include appropriate hardware and software to deal with all

electronic data types, facilities to migrate data as hardware and software become outdated, and extensive resources to recover data from media that are degrading.

The ad hoc SDAC recognizes that, even under the best of conditions, the Institution will have limited resources to manage the Scientific Archives, and therefore it will not be possible for the archives to manage all types of data equally well (or possibly to manage or upgrade some types of data at all). In addition, the ad hoc SDAC cannot predict how technology will evolve or what resources will be available for archiving at any given time in the future. Both will change continually. We therefore make the following general recommendations.

- *Establish a permanent Scientific Data Advisory Committee.* The SDAC would be expected to be an active committee, meeting at least monthly to deal with archiving issues. Its membership should include a representative from each science department, from CIS, from Marine Operations, and from the Deep Submergence Operations Group, as well as the Scientific Archivist and the Data Librarian. The primary SDAC function would be to work with the Scientific Archives, providing breadth of expertise on scientific issues related to data collection, documentation, and archiving, and developing both short-term and long-range perspectives from the scientific viewpoint. SDAC concerns would include the following: a) monitoring evolving technology of data collection and archiving, and advising on adaptations and upgrades needed in the Scientific Archives, b) evaluating and updating priorities for types of scientific data to be archived, c) evaluating space needs for both the Data Library and the Seafloor Samples Laboratory, d) reviewing existing archived data and culling unwanted or unusable data, e) providing advice on which existing data should be migrated, upgraded, or moved to other storage facilities, f) considering other archiving issues, including periodic review of archiving policy and mechanisms. As a group responsible for providing scientific advice and perspectives, the SDAC would report to the Senior Associate Director and Director for Research.

  For the SDAC to fulfill its obligations, committee members will have to devote significant time to their duties, particularly in the first years of SDAC operation. We recommend that WHOI pay the committee member's salaries for the time that they devote to committee work. Appropriate levels of support are suggested to be up to 60 hours per year for SDAC members and up to 120 hours per year for the SDAC chairman. Thus the Institution salary commitment would amount to about 4 man-months per annum.

- *Establish standards for digital data and encourage their use.* The SDAC and the Scientific Archivist should explore the use of standards for field collection, documentation, and archival submission of digital data, and they should request that these standards be utilized by PIs whenever possible. The SDAC and Scientific Archivist should also work with the Data Librarian to establish standards to be used within the archives (e.g., for metadata). We note that acceptance by the WHOI scientific community of promulgated standards is likely to be a hit-or-miss proposition. However, standards have been applied successfully in programs such as WOCE (see Appendix IV). Even partial compliance by PIs will make archiving tasks more manageable.

- *Identify threatened, high-priority data sets and pursue their rescue.* The threat comes from two directions: loss of software/hardware systems to read and utilize digital data,

and deterioration of media. The former is very difficult to predict. Part of the duties of the Scientific Archivist and the SDAC should be to monitor changing technology, evaluate important digital data sets that are becoming threatened, and seek the resources needed to migrate the data. Unfortunately, this is a reactive, unpredictable mode of operation, but we see no viable alternative at this time. Two factors should be considered during this review process. First, data being considered for rescue should rank highly in terms of their potential use to, and historical use by, the scientific community. From the perspective of both inherent scientific value and available resources, it simply may not be feasible to rescue some data. Second, consideration should be given to rescuing some kinds of data (when and if needed) through the services of outside contractors. There are numerous commercial enterprises that specialize in this kind of work. Deterioration of media should be monitored and predicted by the staff of the Institution Archive (see Appendix III).

- *Hire a Cataloger, a Clerical Assistant, and a Mixed-Media Preservationist.* These individuals are needed to manage and rescue existing data within the Institution Archive, as well as to manage future acquisitions and make them accessible to the scientific and educational communities. Details of these positions are given in Section VII.B.

- *Establish and use an efficient metadata system.* Metadata are essential for managing the Scientific Archives. They should be digitally accessible and provide basic information about data type, data quality, where and when data were collected, who collected the data, and where and in what form the data may be accessed. Consideration should be given to establishing a WWW URN for the Data Library metadata. The metadata should be accessible via the World Wide Web using WWW technology or any future similar technology that is easily available on most Internet-connected workstations. One possible approach is to link archived data and metadata to related bibliographic listings, and to use existing, currently available bibliographic search methods for access. WHOI can use the library's on-line system Mariner to link to distributed databases that have both metadata and primary data. Ideally, all the metadata will eventually be in the Mariner system for easier search and retrieval.

  In the case of the Seafloor Samples Laboratory, digital metadata and data bases already exist and they are maintained in MUDDIE and SEDCORE for local access (see Appendix III). These data are also routinely provided to NGDC, where they are web-accessible. This eliminates the need for WHOI to maintain a web site for the WHOI seafloor samples data. The NGDC web site should be linked from the web pages of the Data Library and the Seafloor Samples Laboratory.

  We recommend the following general priority order for establishing and operating the Data Library metadata system: a) develop the system and web presence, keeping flexibility to accommodate the wide variety of scientific data that are or can be archived at the Institution, b) incorporate the Alvin Archive as the initial metadata set, c) incorporate other archived data following the priority order given in Appendix V. This process should be guided by the SDAC. As appropriate during this development, other distributed archives should be web-linked to the Data Library web page.

- *Consider developing a mechanism similar to the Independent Study Awards to award internal, proposal-based grants dedicated to enhancing the Scientific Archives.* Such

grants would raise the awareness of the scientific and technical staff about the importance of the archives, involve expert staff more directly in important archiving projects, and stimulate innovative solutions to archiving problems. The grants could cover topics such as creation or maintenance of important WHOI databases, development of key software tools, and development and maintenance of archive-linked web pages. Proposals would be merit-reviewed in a manner like the present ISA proposals.

**2. Data volume**

In addition to a wide diversity of data, the Scientific Archives must also deal with large data volumes (physical and digital volumes). This creates both space issues and management issues (e.g., hardware and software). The space currently taken up by individual kinds of scientific and other data in the Data Library and in the Seafloor Samples Laboratory is summarized in Appendix III. The projected volumes of annual additions are listed in Appendix V and discussed in Section VI. We make the following recommendations to deal with issues of data volume.

- *Follow a priority list in acquiring and managing scientific data.* We have included in Appendix V our prioritized list (by department) of which data we consider most important to archive. This list should be followed by the archives in allocating available resources. The SDAC should provide guidance to the Scientific Archives on sub-priorities within these lists and should periodically update the priority lists.

- *Utilize other archives and archive-like resources to the extent possible.* These resources are outlined in approximate priority order in Section II.I. The Scientific Archives should develop and maintain web-accessible metadata for WHOI data that are archived using these other resources. Irrespective of whether certain data are archived using the other resources, there may be some of these data that we also want to have on-site for the convenience of WHOI scientists or the WHOI Information Office. Such decisions should be made by the SDAC in consultation with the interested party.

- *Consider using mass-store systems for large volumes of high-priority digital data.* Some kinds of high-volume digital data may be best stored in, and accessed from, a mass-storage system. Examples include phase-difference bathymetry and sidescan-sonar data from the DSL-120, electronic still camera images from NDSF operations, and SeaBeam multibeam bathymetry and sidescan-sonar records. Two mass-store systems currently are operational within the Institution. One is used primarily for storage of large volumes of marine seismology data, and the other is designated for DSOG data. The latter currently is underutilized, and we recommend that the SDAC investigate whether this resource should be developed for archive purposes.

Mass storage facilities might also be subcontracted out to a business specializing in this area. Or WHOI might wish to form a strategic alliance with a company that specializes in managing and archiving data.

Before any decisions are made about mass-storage facilities, there must be a thorough evaluation of the need for such a system, the kinds of data that it would archive, and the long-term costs of maintaining it. Any data to be archived in such a system would also have to meet specific criteria for inclusion, as specified by the SDAC.

• *Investigate the feasibility of using automated tracking systems for data.* Efficient management of large volumes of scientific data may be enhanced by taking advantage of automated tracking technology. For example, bar coding of data when it is first acquired at sea could facilitate subsequent tracking and data management.

**3. Unwanted data contained within the Scientific Archives**

Some kinds of scientific data will lose value over time because they are replaced (e.g., paper records may be digitized), because they are superseded by better data, or because they have degraded beyond recovery. In addition, there are extensive miscellaneous scientific records that been accepted into the Scientific Archives but that have never been sorted through to determine whether they have any ultimate value. There must be a mechanism to cull unwanted and worthless data from the archives.

• *We recommend that the SDAC, in consultation with appropriate interested parties, routinely review the archive collection and cull unwanted data.* The Scientific Archivist should present to the SDAC a list of holdings to be reviewed, prioritized in order of importance (e.g., starting with those holdings which occupy the most space and also appear to have the least scientific value). It is essential that any decisions to discard scientific data be made on the basis of the fullest possible information. Thus, the SDAC should conduct the review with the scientist who deposited the data, or, if the scientist is not available, should consult with another knowledgeable scientist who has expertise in determining the value of the data. Scientific data should *never* be discarded without proper scientific review.

Because such evaluation could involve a significant amount of a scientist's time (e.g., hours to days, or more, depending on the data being surveyed), there must be an incentive for the scientist to participate. We recommend that WHOI cover the scientist's salary for such tasks, within pre-determined limits established by the SDAC in consultation with the Senior Associate Director and Director for Research.

**C. Accessing Data from the Scientific Archives**

**1. Who can obtain data from the Archives?**

Most of the scientific data in the Scientific Archives is obtained with public funding, and therefore it should be accessible to all interested parties. If a question arises about whether access is appropriate, the Scientific Archivist should consult with the Library Director, and, if necessary, the Directorate and the SDAC.

**2. How can metadata (information about data) be accessed?**

This is the essential issue in regard to facilitating access to data in the Scientific Archives. Our recommendation, as outlined in Section B.1 above, is that the Data Library work toward developing an efficient metadata system that is web-accessible. The metadata should include information about where and how data can be accessed from the Scientific Archives.

**3. What data can be obtained?**

At a practical level, data will always be provided "as is", i.e., it should not be the function of the archives to process or upgrade data in response to a data request by a user (the Scientific Archivist, however, may be able to recommend resources within or outside the Institution to assist with reading, processing, or upgrading data).

Users can access data in one of three ways. a) They can visit the Scientific Archives and study the data there. Handling of some fragile data may require supervision by the archives staff. 2) WHOI staff and members of the local scientific community can sign out certain kinds of original scientific data, with a guarantee to return the data within a specified period (i.e., the archives serve as a library). The Scientific Archivist may restrict such use depending on the value of the data and the qualifications of the user. 3) Data may be reproduced under the supervision of the Scientific Archivist, with the copy provided to the user. The user will pay for staff time and reproduction costs in fulfilling the request.

As a matter of Institution policy, original data should not be removed from the Institution by a user unless the user guarantees safe return of the data. Decisions on such requests should be made by consultation among the Scientific Archivist, the SDAC, the Library Director, and the Directorate.

**4. Restrictions on use of archived data**

Data deposited with the Scientific Archives will be proprietary to the Principal Investigator for a period that accords with the policy of agencies funding the research and with Institution archiving policy. This period is normally two years (see Section IV.A), after which the data generally should be made available to other users.

There are legal questions about ownership of data, rights to commercial use, and so forth that are beyond the scope of the ad hoc SDAC deliberations. However, we agree on the following general principles. In most cases, data accepted into the Scientific Archives become the property of the Institution (exceptions include, e.g., data loaned to the Institution). Data which were obtained with public funding should be freely available to the public for the purposes of scientific research, historical research, and education. The Institution may restrict access and use of data obtained with private funding. The Institution should retain commercial rights to data within the Scientific Archives. Any funds derived from commercial use of archived data should accrue both to the PI who archived the data and to the Institution. The possibility that funds may accrue to a PI is perhaps the most significant incentive that WHOI can offer to encourage PIs to archive their scientific data. Funds going to the PI should be used to promote his/her research; the Institution may want to restrict this benefit to only WHOI scientists who are actively conducting research. Funds accruing to the Institution should be entirely devoted to maintaining and upgrading the Scientific Archives.

## VI. SPACE NEEDS

The ad hoc SDAC reviewed space needs for archiving facilities. Both the Data Library and the Seafloor Samples Laboratory are full, and a synopsis of immediate space needs is given below. We also attempted to project future space needs for both facilities. This is difficult because, at least in the case of the Data Library, required space will

depend in large part on the archiving policy that the Institution adopts. Our projections are given below, together with underlying assumptions on which the projections are based.

**A. Data Library**

**1. Current needs**

*Space for personnel and users.* Office space and a suitable area for data processing and patron use is currently needed in the Data Library. The clerical staff share the same office space and do not have permanent workstation assignments. Patrons need an area, including a substantial work table, for reviewing data, maps, and charts. Additional space is needed for the new, enhanced, networkable microfilm reader/printer.

*Data and historic archive storage area.* Compact shelving for the Film and Tape Vault is needed. Materials (e.g., Alvin films) overflow into the archive vault, the main stacks area, and, most recently, the Map Storage Room. There is no vault storage space for incoming media, and films are temporarily being kept in the Data Librarian's office.

The archive vault is full. It contains the Directors' files up to the 1980's which are stored in file and microfilm cabinets (a poor use of space). Later materials are now stored in the main stack area, but there is little room even there or in the back storage room for incoming accessions. The 100 scientific instruments on loan to the Naval Museum will be returning in a few years and there is *no* room for those items.

*Summary of immediate space requirements.*

| | |
|---|---|
| Estimated area needed for staff: | 600 square feet |
| Estimated area needed for users: | 200 square feet |
| Estimated area needed for data, film, and tapes (compact shelving): | 400 square feet |
| Total immediate space needs: | 1200 square feet |

**2. Freeing up currently used space**

There is potential for making some space available by discarding unwanted or unusable data. Some materials also may eventually be sent to the Boston Library Consortium's new off-site storage facility, but this facility will not be available for at least the next 3-5 years. Determining the amount of space that can be freed up by these means will require in-depth assessment of current holdings by archive staff and review by the SDAC.

Some of the space problem can be alleviated if the data/archive section acquires a system of compact shelving, which will use current space more efficiently. A new Spacesaver system would allow for 8 more banks of shelving, which would add an additional 1344 linear feet of storage space (432 square feet).

**3. Future space needs**

*Document/technical report area.* The recently installed Spacesaver shelving will probably accommodate 5 to 10 years of growth in documents and technical reports. This is based on an estimated 6 linear feet of new material accessioned yearly, taking into account further weeding and collection shifting.

*Data/archive section.* The archive currently takes in an average of 150 linear feet (36 square feet) of data annually in all formats (e.g., film, tapes, video, echo-sounding records, CDs, computer diskettes, photographs, slides, scientific papers); much of the data is from researchers currently active at WHOI (e.g., B. Lange, J. Howland, R. Stephen). Included are materials from the administration and the Education Office, which can have huge fluctuations in volume from one year to the next.

Several factors could better than double this acquisition rate in coming years. A proactive campaign to consistently gather newly collected, primary scientific data certainly will increase the amount of routine acquisitions in the archive, and we have made estimates below. Acquisition of *existing* scientific data that currently resides with PIs would greatly increase the holdings; we have *not* attempted to estimate the volume of this material. With the retirement of some notable scientists (e.g., R. Ballard), the archive could also be overwhelmed with scientific data and papers that must be kept for future scholars. In addition, the Institution's seventieth anniversary is coming and the seventy fifth is close behind. This will increase awareness of the value of our collections and may lead to increased accessions. The historic archives certainly will be used heavily by staff, scholars, and historians, and this will require suitable work space.

*Personnel space.* With accelerated acquisitions that would result from promulgation of a new archiving policy and establishment of mechanisms to implement it, office space will be needed for the Scientific Archivist and other clerical staff.

*Climate control.* Space needs cannot be addressed without discussion of the need for climate control for all the materials that are stored for future generations. Ideally, temperature- and humidity-controlled space should be made available to allow for the safety of archived materials. Much of the current deterioration (see, e.g., Section I.B in Appendix III) is caused by lack of humidity and temperature controls.

*Centralization of archived data.* The SDAC feels strongly that the Data Library should remain centralized. Impending crisis management may require some short-term diffusion of data, but this diffusion should be remedied as quickly as possible and should *not* become long-term policy.

*Projections.* We have projected future space needs for the Data Library, based on estimated acquisition of new data in individual departments (Appendix V), current annual acquisitions, space-saving measures, and other factors. We emphasize that these are only rough estimates. Improving the estimates significantly will require that experience be gained under an established archiving policy.

Table 1. Annual space required to house *new* influx of Priority 1 data plus *current* influx.

| Data source | Cubic feet | Square feet* |
|---|---|---|
| Geology & Geophysics | 40 | 7 |
| Research Vessels | 32 | 5 |
| Alvin | 39 | 7 |
| Other Deep Submergence | 36 | 6 |
| Biology | 19 | 3 |

| | | |
|---|---|---|
| Marine Chem. & Geochem. | 72 | 12 |
| Physical Oceanography | 0 | 0 |
| AOPE | 2 | 1 |
| Other scientific** | 60 | 10 |
| Other Institution† | 60 | 10 |
| Total | 360 | 61 |

Table 2. Annual space required to house *new* influx of Priority 2 data only.

| Data source | Cubic feet | Square feet* |
|---|---|---|
| Geology & Geophysics | 3 | 1 |
| Research Vessels | 0 | 0 |
| Alvin | 0 | 0 |
| Other Deep Submergence | 0 | 0 |
| Biology | 12 | 2 |
| Marine Chem. & Geochem. | 0 | 0 |
| Physical Oceanography | 10 | 2 |
| AOPE | 2 | 1 |
| Total | 27 | 6 |

Tables 3a and 3b. Projections of *present* and *future* space needs.

| Item | 5 years* | 10 years* | 20 years* |
|---|---|---|---|
| Priority 1 data only | 305 sq.ft. | 610 sq.ft. | 1220 sq.ft. |
| Existing data: shortfall | 400 | 400 | 400 |
| Staff space | 800 | 800 | 800 |
| User space | 600 | 600 | 600 |
| Minus compact shelving†† | 432 | 432 | 432 |
| Total | 1673 sq.ft. | 1978 sq.ft. | 2588 sq.ft. |

| Item | 5 years* | 10 years* | 20 years* |
|---|---|---|---|
| Priority 1 + Priority 2 | 335 sq.ft. | 670 sq.ft. | 1340 sq.ft. |
| Existing data: shortfall | 400 | 400 | 400 |
| Staff space | 800 | 800 | 800 |
| User space | 600 | 600 | 600 |
| Minus compact shelving†† | 432 | 432 | 432 |
| Total | 1703 sq.ft. | 2038 sq.ft. | 2708 sq.ft. |

NOTE: Active solicitation of *existing* data from scientists within the Institution and incorporation into the Data Library could greatly increase these space requirements over the next 5-10 years.

*Assumes compact shelving is used in any *new* space.
**Current average annual inflow of scientific materials (data, scientist papers, etc.).
†Current average annual inflow of materials from the Directorate and Corporation.
††Space savings if compact shelving is used to replace *existing* shelving.

**B. Seafloor Samples Laboratory**

**1. Current needs**

The Seafloor Samples Laboratory is full except for about 500 square feet of floor space. This area is needed as working room to temporarily store, sort, label, and catalog arriving samples, and to box or D-tube them for storage in shelves and racks.

**2. Future space needs**

Sample acquisitions in the Seafloor Samples Laboratory have been fairly steady, and this trend is expected to continue. In terms of square footage, annual additions of samples amount to about 120 square feet (2.5% of currently available space of 4850 sq.ft.). As an example of acquisitions, the Seafloor Samples Laboratory expects in 1998 to receive 150-200 piston cores from three coring cruises, as well as rocks from up to 100 dredges. This is fairly typical for annual additions.

*Projections of future space needs*

| Data | 5 years | 10 years | 20 years |
|---|---|---|---|
| Sample storage | 600 sq.ft. | 1200 sq.ft. | 2400 sq.ft. |

The SDAC feels strongly that the Seafloor Samples Laboratory should remain centralized and that its collection and resources should *not* be distributed throughout the Institution.

## VII. SUMMARY OF OTHER PROJECTED NEEDS

We here summarize our estimates for the needs, other than space, for archiving of scientific data over the next 5-20 years.

### A. Scientific Archivist

The ad hoc SDAC identified the need for a Scientific Archivist to track WHOI scientific data and to accomplish the functions described below.

*Duties*: This individual would be responsible for identifying, tracking, and acquiring new scientific data collected by WHOI investigators and WHOI research vessels and vehicles. The Scientific Archivist would work closely with Marine Operations, individual departments, and PIs to identify data to be collected, promulgate Institution archiving policies, make tools available for data collection and educate users on their application, acquire Science Log Books, Cruise Reports, and other data documentation from cruises on WHOI ships, and acquire and archive scientific data. The Scientific Archivist would also work with the scientific and technical staff to document and archive existing scientific data held by investigators within the Institution. The Scientific Archivist would interact with the Data Librarian, the SDAC, CIS, and other archives staff to develop, implement, and maintain software tools that facilitate data collection, data management, data migration, WWW access, and data distribution. The Scientific Archivist would be a facilitator who is backed by policies (WHOI, UNOLS, funding agencies, other national data facilities) that will help his/her efforts in assisting the scientific staff to comply with standards and reporting practices. The Scientific Archivist would have the direct backing of, and report to, the Senior Associate Director and Director for Research.

*Qualifications.* A strong scientific background is essential and it must include expertise with collection, management, value of scientific (preferably oceanographic) data. Bachelor's degree in oceanography, geology/geophysics, or closely related field, and Master's degree in library science or information-systems management. Or Master's degree in one of the above scientific fields, with extensive experience in data management, including familiarity with standards and guidelines for indexing and cataloging. Strong computer skills and familiarity with wide-ranging computing platforms are essential. Excellent interpersonal and communication skills are imperative. *Highly desirable.* Experience on oceanographic cruises. Understanding of distributed network solutions for data dissemination.

*Comment.* This will be a full time position. The position of Scientific Archivist differs from and complements the current Data Librarian position. The Data Librarian is the managerial position in the Institution Archive, which includes responsibilities related to Human Resources, budgets, planning, task assignments, and overall operation of the facility. A much needed Long Range Plan for the facility will be initiated by the Data Librarian, and the Scientific Archivist will be part of the team that works on this document. The Scientific Archivist will be responsible for proactive acquisition of data from individual PIs and departments from the inception of scientific programs, as well as for acquisition of existing data. The Scientific Archivist will serve as a liaison to the scientific and technical staff who have major collections that need to be assimilated into

the Scientific Archives. While the Data Librarian will oversee the development of the WWW site for dissemination of information, the Scientific Archivist will prepare and analyze scientific data for inclusion in this site. The Scientific Archivist will work primarily in the Data Library but will also interact routinely with the Seafloor Samples Laboratory.

**B. Additional Archive Staff Requirements**

**1. First priority**

*Cataloger - Information Systems Associate I* - To oversee book, atlas, and technical report cataloging, perform necessary original cataloging for unique holdings, and monitor copy cataloging by clerical staff. The need for additional cataloging efforts already exists because there is a current backlog of uncataloged acquisitions, and this need will increase with acquisition of new materials.

*Clerical Assistance - Information Systems Assistant II* - To provide assistance in projects such as copy cataloging, processing incoming reports and serials, maintaining ship's data bases, and filing. Must be computer-literate and facile with network and Internet applications.

**2. Second priority**

*Mixed-Media Preservationist - Information Systems Assistant II or III* - To identify and rescue deteriorating film, tapes, fiche, glass slides, etc.; to scan and digitize threatened media; to clean and duplicate film. Graphics and computer background necessary.

**C. Permanent SDAC**

A permanent SDAC is a critical element to implementing and maintaining a viable archiving policy, and committee members will have a significant work load. The Institution should budget 4 man-months per annum to support the work of this committee.

**D. Support for Scientific and Technical Staff**

Culling unwanted or unusable data from the Scientific Archives is essential in order to make space available for important data and to minimize the rate at which new space will be required as the data collection grows. Experts from the Scientific and Technical Staff will be needed to help the SDAC and Scientific Archivist cull this data intelligently, and they are not likely to invest their time without incentive. We recommend that the Institution initially budget 1 man-month of salary support per year for staff to help cull data; this figure can be revised as we gain experience on levels of effort needed.

**E. Web-Accessible Data System for the Institution Archive**

The Data Librarian will begin to implement a web-accessible system for accessing information (metadata) about the currently available (i.e., non-restricted) data contained in the Scientific Archives. The system will also include links to other web-accessible data sites, both inside and outside WHOI. Eventually, actual digital data may be made available for access.

Initial costs for the web-site development will be for a UNIX workstation and appropriate software and hardware components that will allow the Data Librarian to begin site development. The workstation should be scaleable to accommodate future growth as the database grows. Priority decisions will need to be made on inclusion of data, but NDSF data should take initial precedence and be the starting focus of the web site.

Estimated system hardware and software costs: $10,000

Estimated system development cost, CIS: $10,000

Future development, maintenance, and upgrade costs will depend primarily on volumes and formats of data acquired, and on evolving developments in technology.

**F. National Deep Submergence Facility**

The ad hoc SDAC proposes that NDSF data be acquired and archived as summarized in Appendix VII. This will require some adjustments in the way that data are currently recorded, distributed and archived.

*Alvin video.* Currently, only one master video input is recorded in the DSV. This is duplicated on the ship, with the duplicate going to the scientific party and the original to the Alvin Archive. We propose that two master video inputs be recorded in the DSV. This would require that another tape deck (already available) be routinely mounted in the DSV, and that a second set of tapes be duplicated on the ship. Additional costs would be incurred for one set of master and duplicate Hi8 tapes, amounting to ca. $5000 per year. Both master tapes would be archived, and both duplicate tapes would go to the scientific party.

*Jason and Argo II video.* Currently, three master video inputs are recorded, and all of these tapes are given to the scientific party. *None* of the video data are archived. We propose that two master video inputs be recorded, and that two duplicates be automatically recorded on slave recorders. The two master videos will go to the Scientific Archives, and the two duplicates will go to the scientific party. By doing this, Jason and Argo II video will be treated in the same manner as the Alvin video. Additional costs would be incurred for the one-time purchase of a Hi8 video tape deck (ca. $1200) and for one set of Hi8 tapes (ca. $120 per full day of Jason or Argo II survey).

*DSL-120 Sonar*. Currently, raw sonar tapes go to the Scientific Archives and tapes of processed sonar go to the scientific party. A copy of the processed sonar should also go to the Scientific Archives. Additional costs would be incurred for the one-time purchase of an extra high-density Exabyte tape drive (ca. $2100) and for one set of Exabyte tapes (ca. $50 per full day of DSL-120 sonar survey).

## VIII. DATA ARCHIVING AND SCIENTIFIC JOURNALS

The ad hoc SDAC considered two issues concerning the interplay of data archiving and policies of scientific journals. It was pointed out that scientific journals will not accept or publish research papers dealing with the Human Genome Project until it is demonstrated that the associated research data are archived and accessible to the scientific community. We briefly considered this as a model to compel proper archiving of scientific data, but we dismissed it as unworkable. It would be impossible to apply this model to the extremely broad spectrum of oceanographic data, and it would induce resentment among investigators rather than promote the advantages of archiving.

The other issue considered was the disposition of auxiliary data (e.g., long data tables) that sometimes supplement published research articles. Most journals have a mechanism to accept, "archive", and make such data available to interested parties. However, WHOI investigators should be made aware that the Scientific Archives will archive the data if

necessary, and the investigator can indicate in his/her publication that the data can be obtained from the Scientific Archives upon request.

## APPENDIX I

## SUMMARY OF AD HOC SDAC MEETINGS

**Meeting dates and major issues discussed.**

**11 September 1997** - Committee charge, scope of problem, and definitions. Existing policies and archiving practices. General work plan and schedule.

**2 October 1997** - Visit and survey of the Data Library. Review and discussion of types of data acquired in each department. Definition of specific problems and assignment of committee members to work on solutions in each area.

**20 October 1997** - Discussion and agreement on generic archiving policy. Mechanisms, needs, implementation. Problem discussion (becoming aware of and tracking data; acquiring data for the archives).

**13 November 1997** - Review of facilities (e.g., NOSAMS) and Seafloor Samples Laboratory. Problem discussion (tools, incentives, mechanisms to acquire data for archives; data management; distributing data). Space issues.

**24 November 1997** - Archive policy for NDSF. Review and discussion of draft outline for SDAC report. Writing assignments.

**15 December 1997** - Discussion of outside loan of MBT collection. Prioritization of data to be archived. Mechanisms to cull data from the archives. Review of writing assignments.

**12 January 1998** - Review, discussion, corrections on draft SDAC report.

## APPENDIX II

## COMMENTS FROM THE WHOI SCIENTIFIC AND TECHNICAL STAFF

**Physical Oceanography Department**

- Several newer technologies (instruments) have been developed in the past few years which do not have formal archiving yet. There is no rule regarding the sharing of these data, nor is there a structure in place to handle their archiving or dissemination.
- Data that have been acquired under the auspices of one of the large super-programs (such as WOCE) are reasonably well organized. Yet, neither the historical nor future data will be similarly accessible, unless effort is directed at incorporating those other data into a unified domain.
- Concomitant with the growth of WOCE, a drop in submission to the NODC archives has occurred. Submission is largely unenforced. Frequently cited by various PIs as reasons for this unfortunate state: " There is no easy way to do this. The tools do not exist to facilitate submission of this data. No one has come banging on my door, so why worry about it? I can't be bothered with NODC--they charge me money to get the data out again."
- With the fading of centralized computing (VAX), the databases for CTD, hydrography (bottle samples), current meters, and buoy data have dissolved. There is no structure or group in place to oversee the management of these data, and no funding to perform such tasks.

**Biology Department**

- Archiving services: It would be useful to have on-line access for 5 years, then archive at a federal facility, possibly the funding agency (ad hoc SDAC note: funding agencies generally have no such capability; this comment highlights the lack of general knowledge that many investigators have about data archiving).
- Archiving services: Yes, [keep] at WHOI. [A] national [archive] is worth considering as long as WHOI researchers have priority and guaranteed access.
- Archiving services: Some data will be easier to archive and be used more often than other types.
- Data retrieval services: For recent data, Internet access would be useful. Eventually data might be migrated to, e.g., CD ROM or other media.
- Data Retrieval Services: Imperative that there be controls over the quality and distribution of images. Also, there needs to be some way of guaranteeing acknowledgment of the data source. Lastly, there needs to be some way of limiting data access to only qualified researchers.
- Policy: We'll probably need flexibility.
- Policy: Regarding the time-schedule for deposit of data with the Scientific Archives, the Principal Investigator should be able to specify this schedule (ad hoc SDAC note: the ad hoc SDAC disagrees).
- I've never sent my data to a national data center. I suppose I should, but my records are so short and so limited spatially that I didn't think they would be of much use to others.

**Geology and Geophysics Department**

- Overall, G&G staff were very positive about the idea that WHOI should have a proactive archiving policy and that scientific data should be archived. Some felt very strongly that documenting and archiving scientific data should be an essential function of the Institution. These included individuals who expressed frustration about data that should have been archived and accessible for study but that has disappeared and appears to be irrecoverable.
- Most, when asked, admitted that they have done a poor job of seeing to it that their own data are archived either at WHOI or with a national data center. The common explanations were that it was just another task that they didn't have time or resources for, and that no one seemed to care, so why bother?
- Some were worried that their data might be "lost to science" if not archived, and they welcomed the idea of working with an archivist to document and archive their data. Some noted that their data were of such small volume, or so arcane, as to be of little value to the archives.

**Applied Ocean Physics and Engineering Department**

- No specific comments were received from AOPE.

**Marine Chemistry and Geochemistry Department**

The scientists in the Marine Chemistry and Geochemistry Department have diverse research interests that involve a number of different data types. These include remote sensing, various hydrographic data, time series studies of chemical fluxes, and they also include many physical sample types (e.g., sediment traps, deep sea sediments, sediment pore waters, oceanic volcanic rocks, terrestrial glacial deposits, and hydrothermal deposits). Laboratory studies of these samples utilizes diverse techniques and instrumentation (e.g. mass spectrometers, gas chromatographs, radioactive decay counters). During several informal discussions some general concerns were voiced:

- There was a general consensus that data archiving is a valuable activity, but there was some concern that scarce unrestricted funds should not be diverted from supporting scientific research (Note: the ad hoc SDAC considers documentation and preservation of data for future analysis to be an essential part of scientific research).
- Some of the older (hydrographic) data residing here at WHOI are still very valuable even though they are not in digital form and therefore are not readily accessible. Considering the value of such data, it was recommended that data be archived for at least 25 years.
- Data archiving, coupled with library services, should be maintained at a high level to support scientific research. This will be challenging because of increasing costs of journal subscriptions and because of uncertainties about the long-term viability of, e.g., on-line journal subscriptions.
- Enough resources should be made available, particularly in the form of a dedicated WHOI data archivist, to facilitate proper archiving on the part of the scientific staff.

## APPENDIX III

## SUMMARY STATUS OF THE SCIENTIFIC ARCHIVES

### I. DATA LIBRARY

**A. Background.**

In 1956, J. B. Hersey, Department Chairman of G&G, established the Data Library. He recognized the need to organize geological and geophysical data into a workable collection, as well as a need to staff the library with a knowledgeable person. Historically, the main purpose of this archive has been to provide WHOI scientists and students with a working data collection that systematically preserves scientifically useful and historically significant data sets. The collection has broadened over the years to include material from all the WHOI scientific departments, gathered from all the world's oceans as well as from seas, lakes, coastal zones, and rivers. These data are in a wide variety of formats and types including written logs and notes, plots, films, video-tapes, magnetic tapes, CD ROMs and other digital formats in a variety of storage systems. Data in the Data Library belong to the Institution unless otherwise noted by a legal document. PIs have proprietary rights to their data for the period of time noted in their grant awards, and no data is released during this period without specific permission of the PI. There are currently no consistent mechanisms for acquiring scientific data for the Data Library. Scientists' data and papers usually arrive in an unsystematic fashion (if at all) when the PI retires or dies.

The Data Library also maintains important specialized collections. Among these is the Alvin Archive, the bulk of which is underwater imagery in the form of 35mm slides and video tapes. The films are stored in fireproof vaults and the video cassettes are stored in a magnetic-tape storage vault. Alvin and NDSF data are the most consistently submitted and archived scientific data in the Institution.

Another specialized collection is the Scientific Instrument collection. Its contents is summarized among the following pages.

**B. Physical Condition and Accessibility of the Collection**

**1. The Alvin rescue project**

The first phase of this project involved cleaning and duplicating Alvin films up to Dive 1500. All outdated film stock was transferred to the new process E6. The old, original film (E4 process) was retained. During the second phase of the project, the Graphics Department is up to Dive 1800 and has identified another 250 images for possible publication in a second volume of "Alvin's Greatest Hits". These images will be reviewed by Dan Fornari and Shelley Lauzon. All the Alvin dives have been cataloged and are in the library's Mariner system. All 300 CD images from the first volume of "Alvin's Greatest Hits" have been scanned and linked to the library catalog; they are available over the Internet for viewing at low resolution but are certainly not of publishable quality. All are identified as copyrighted by WHOI.

**2. Film vault**

The Film Vault contains 1) 9-track magnetic tapes (800/1600 BPI) of data mostly from the P.O. Department (e.g., CTD data) and the G&G Department (gravity and old DDL data), 2) 8mm, 16mm and 28mm motion picture film (historical WHOI footage), 3) 35mm slides, 4) VHS and Beta tapes, 5) audio tapes. 6) color and B/W photographs, and 7) Alvin film footage. All of this material is deteriorating. None of the 9-track tape can be read at WHOI. It will need to be sent to a service center in Boston if the data are to be recovered and migrated. This issue should be reviewed by the permanent SDAC.

**3. Paper seismic reflection records**

Many of the paper seismic-reflection and echo-sounding profiles acquired during the era of the WHOI Precision Graphic Recorder are fading badly and may eventually become useless. The full status, scientific value, and ultimate disposition of these PGR records, including determination of which records are or should be microfilmed, remains to be assessed. This assessment should be given a relatively high priority because the records occupy significant physical space in the archive.

**C. Physical Space Occupied by the Collection**

The attached sheets summarize the physical space occupied by current holdings in the Data Library and Institution Archive. They provide a useful perspective on relative abundances of various kinds of data. By far the most voluminous component is echo-sounding profiles. Evaluation of the value and need for ready access to these data should be a high priority; these, and possibly other data, might be culled or moved elsewhere (e.g., the Boston Library Consortium's new off-site storage facility).

**D. Services Provided By WHOI Staff in the Data Library**

• General reference service for all aspects of ocean science, ocean engineering, and geophysical data obtained in WHOI research, administrative, and education activities. Queries may be received via phone, mail, email, and walk-in visits.
• Duplication of requested materials by photocopying, photography, or tape duplicating. This includes copying Alvin film and sending it to the Principal Investigator.
• Maintenance of cruise data and ships' logs, records, and tracks.
• Check in and cataloging of periodicals.
• Exchange of WHOI technical reports.
• Generation of reports, publications, and finding aids (Manuscript Collection series; Archival Collection series; ONR publications; Bibliography of Technical Reports).
• Acquisition, cataloging, maintenance, preservation, and providing access to the following resources: WHOI data which are in numerous formats in a variety of storage systems; technical reports and documents from WHOI, from other universities, and from government, non-profit, and industrial research agencies; corporate reports; cruise logs and records; scientific instruments; scientific and administrative papers and records; Institution photographs; Alvin Archive; monographs and journals; films, videos and mixed media.

**E. Current and Future Major Projects (in order of descending priority)**

- Rescue deteriorating material and media of value to data collection.
- Obtain scientific input on relative value of identified items.
- Obtain scientific input on material that can be culled and discarded.
- Clean and, if necessary, duplicate all NDSF films.
- Make metadata about NDSF data available on the web.
- Recall overdue items (atlases, maps, charts, data, photos, etc.).
- Database entry and logging of Alvin video tapes. Shelley Lauzon specifies films to be dealt with.
- Clean old films. We have a new film-cleaning machine that can clean any films up to 70 mm.
- Thousands of bottom photographs need preservation, inventorying, etc.

- Ship's papers, now stored in file cabinets, need to be processed (preserved, re-foldered, boxed, listed).
- Scientists' papers need processing (e.g., Ketchum, Sears, Iselin, Redfield, Bumpus, etc.).
- Re-sleeve thousands of negatives out of glassine sleeves, including Jan Hahn negatives.
- Instrument collection: cleaning; assigning new locations for every single one. Frank Taylor, a volunteer, has been the sole caretaker of this project, and additional help is needed.
- Many boxes of accessioned data had been put in the map storage room; these need to be re-boxed into standard-size cartons and stored in the central shelving units. Much of this data came from Bill Lange and includes Alvin film. These materials need listing and indexing.
- Anticipate sending materials to the Boston Library Consortium's new off-site storage facility and mark items for analysis. Need to determine what to send, inventory and box materials, and bar-code items. ETA of this new facility is 5 years.

**F. Future Services We Should Provide**

- Complete electronic records of Data Library contents for on-line retrieval of information.
- Web links to original data sets either locally or at other national facilities.
- Further on-line access to WHOI data and map collection.
- Electronic delivery of information via the Intranet and Internet.

**G. Survey of the Contents of the Data Library and Institution Archive, conducted by Margot Brown Garritt on 15 December 1997.**

## II. SEAFLOOR SAMPLES LABORATORY

### A. Background (adapted from the WHOI SSL web page)

The WHOI Seafloor Samples Laboratory is a world-class facility that houses more than 13,000 geological samples, largely from the seafloor of the world's oceans. The sample inventory includes long, stratified sediment cores, rock dredges, and grab samples recovered by conventional seafloor sampling equipment and by DSV Alvin. Detailed descriptions exist for all samples in the collection. To date, twelve volumes of descriptive material have been published and distributed to the marine science community via libraries and research centers around the world.

Sediment cores are initially described with respect to color, grain size, unit contacts, and presence of macroscopic fauna and bioturbation. Lithologies are determined by a combination of visual examination and quantification of biogenic and lithogenic components present in smear slides prepared from each core. In addition to lithologic and microscopic analysis, all incoming cores are photographed when fresh to preserve a visual record of original core condition. High quality 4x5 black-and-white negatives are on file in the laboratory. Prints are made available to the user community on request at nominal cost. Full core x-radiographs may be undertaken when prescribed by individual research interests. Rock samples are numbered, weighed, and named according to composition and texture. Notations are made regarding grain size, mineralogy, phenocrysts, vesicles, amygdules, coatings, weathering, and alteration.

### B. Physical Condition and Accessibility

Cores are stored in sealed D-tubes to prevent drying. Rock and grab samples are stored in rack boxes. Limited support from NSF, ONR, and USGS allows this collection to be maintained at a reasonable level.

Data about each sample exist in digitized form. Data retrieval is possible within the laboratory via the software program MUDDIE. Users may sort the vast database by specifying selected parameters. Although MUDDIE has been kept up to date for "quick and dirty" data searches, more sophisticated software (SEDCORE) developed by Roger Goldsmith is now being used. This is a SQL database search/retrieval system currently accessible only in the Seafloor Samples Laboratory. The sample data are routinely furnished to NGDC, where they are accessible on the WWW at: http://www.ngdc.noaa.gov/mgg/curator/cursrch.HTML

Documentation is lacking for some assemblages of samples that have been deposited in the Seafloor Samples Laboratory by individual scientists. It should be a high-priority task of the Scientific Archivist to contact and work with these scientists on properly documenting the samples.

Subsamples from the SSL collection are distributed around the world at no cost to researchers within the scientific community who express a legitimate interest and need. 14,004 subsamples were distributed in 1985-87, 15,089 subsamples in 1991-93, and 20,525 subsamples in 1994-1996.

### C. Physical Space Occupied by the Collection

The SSL collection is housed in an area of 4850 square feet, which includes 500 square feet of refrigerated storage. The laboratory is now full, except for 500 square feet which is needed as a working/staging area to manage the collection.

Assuming no significant changes in sample acquisition patterns, the seafloor sample collection will continue to grow at about 120 square feet annually. There will soon be no working/staging area available to manage acquisitions, and within about 4 years there will be no space available to house *any* further acquisitions.

**D. Services Provided By WHOI Staff**

The staff of the Seafloor Samples Laboratory documents, describes, manages, and maintains the sample collection. They routinely provide sample descriptions and metadata to NGDC and the global scientific community, and they distribute subsamples in response to requests.

**E. Future Services We Should Provide**

The Seafloor Samples Laboratory does not proactively seek samples or create metadata on samples that are not provided to the laboratory directly by the PI. Thus, samples acquired by WHOI investigators or by WHOI facilities (including Alvin and Jason) may not be documented or archived by the laboratory. The Scientific Archivist, working with the Seafloor Samples Laboratory, should assure that *all* such samples are fully documented, tracked, and if possible acquired for the seafloor sample archives.

## APPENDIX IV

## STATUS OF ARCHIVING STRATEGIES FOR MAJOR PROGRAMS AND FACILITIES CENTERED AT WHOI

### U.S. JGOFS and U.S. GLOBEC Georges Bank Programs

Both the U.S. JGOFS and U.S. GLOBEC Georges Bank Programs use the JGOFS Data Management System to organize, store and serve data (JGOFS Data System Overview, G.R. Flierl, J.K.B. Bishop, D.M. Glover, and S. Paranjpe, URL http://puddle.mit.edu/datasys/jgsys.html). Data are accessed and retrieved via the World Wide Web using any standard browser, such as Netscape or Internet Explorer. This data system enables investigators to share scientific results with colleagues via a consistent and useful interface, independent of computing platform. Some data analysis and plotting tools are available on-line from within this same system. Researchers can download data of interest for further study and analysis.

Data and supporting documentation eventually will be archived to suitable national archiving facilities such as NODC. Also, U.S. JGOFS plans to generate CD ROMs containing the data from their program and distribute these widely.

### WOCE

Data management and quality control are both highly organized functions within the International WOCE Program. Data management is implemented under a distributed system: each measurement technique produces a different data stream, which the data management system brings together to form a single data resource. This system is comprised of several elements in which the flow of data moves from Principal Investigator to Data Assembly Centre (DAC), to Special Analysis Centres (SAC), to users and archive. The Data Information Unit (DIU) is a central web-based source of information (metadata) on the status of WOCE, tracking all data collection, processing and archiving activities, and acting as the primary interface between the WOCE data system and all users.

DACs are managed by scientists, handle assembly and quality control of data sets, and generate data products.

SACs perform data analysis and synthesis functions, including the generation of derived data sets.

The WOCE Archive is distributed by data type and location according to a predetermined World Data Center System. Access to this archive is via the World Wide Web.

Perhaps one of WOCE's greatest legacies will be the unified formats for each type of instrument, to which all international participants must subscribe. Even if it is painful, at least every PI has programs to read and write WOCE format, and this has been a great facilitator to the exchange of data. A second legacy will have been the quality control applied to each dataset -- a significant problem that plagues the earlier historical data. The WOCE system excels also in its management of metadata -- the DIU is a web resource which appears to be well managed and current in tracking data collection and status. The practical deficiencies of the system lie in enforcement of its submission

policy (within 2 years of collection), in a backlog of performing the quality control, and hence in the timeline for making the data publicly available. With respect to the future: technically, WOCE has a beginning and an end. The program is committed to archiving and preserving the WOCE data set in its entirety under the auspices of the World Data Centre (WDC-A) at the NODC. A searchable data retrieval system will be developed at WDC-A and will provide continuity into the future when the DACs and SACs are no longer active. Perhaps this archive will continue as the de facto standard within the international community, by virtue of the amount of infrastructure that is already in place.

**NOSAMS**

The National Ocean Science Accelerator Mass Spectrometer facility maintains an SQL-queriable relational database of accelerator mass spectrometry analyses. Data are recorded directly from the instrumentation to the database, and the database is not currently web-accessible. From an archival perspective, a current deficiency in the way data are recorded is that the provenance of samples (details on sample origin, location, etc.) is not consistently known and documented.

NOSAMS is considering upgrading the database to "archive quality" and making it web-accessible. This would be supported by NSF through normal, incremental funding of the facility. However, there are issues to be resolved before such access can become a reality. One issue is that of instituting a mechanism to obtain provenance information that will make the data scientifically useful to the general community. Another is the question of establishing policy for a proprietary period on the data. With the exception of WOCE analyses, there is currently no NOSAMS policy on this matter. On the way to resolving both issues, there may be some resistance from users of the facility. This is somewhat problematical (but not an insurmountable problem) in that NOSAMS must compete for business with other accelerator mass spectrometry facilities (Livermore, Arizona).

The bottom line is that NOSAMS is amenable to fully documenting and archiving data and making it web-accessible, and this could happen at no cost to the Institution. NOSAMS awaits information on the archival policy that will be adopted by the Institution. The ad hoc SDAC recommends that this issue be pursued with the goal of making fully documented NOSAMS data web-accessible.

## APPENDIX V

## DATA CURRENTLY ACQUIRED ON WHOI SHIPS AND BY WHOI INVESTIGATORS

This appendix attempts to summarize the kinds of scientific data currently being acquired by WHOI seagoing facilities, and by WHOI investigators both on WHOI ships and elsewhere. We have attempted to be complete, but the listings are not exhaustive. Tables list data for each department. Table headings are:

**Priority (by consensus of ad hoc SDAC and staff responses).**

1 - Data that must be archived.

2 - Data that should be archived if resources are available (funds, space, personnel). These are listed in order of generally descending priority.

3 - Derived data that might be archived upon the request of the PI if the data are not available elsewhere. The Scientific Archivist must be knowledgeable about all national data archives that accept this kind of data and may refer the PI to such a repository. Acceptance into the Scientific Archives requires approval by the SDAC.

The order in which data are actually archived may be influenced by availability of resources for particular kinds of data.

**Nature of data.**

Self explanatory.

**Media type.**

Self explanatory.

**Volume (cubic feet).**

Our best estimate of the annual volume of archive space that the data would require.

**How archived.**

The current method by which the data are archived.

**Duration.**

Assessment of how long the data should be held in the Scientific Archives, by consensus of the ad hoc SDAC and staff responses. This assessment is based on current conditions and could change with changes in e.g., technology, policies of other data centers, etc.

**Notes on SeaBeam swath bathymetry collected under the auspices of Marine Operations (adapted from http://mbdata.whoi.edu/mbdata.html)**

WHOI operates SeaBeam 2100/12 multibeam bathymetric sonar systems on R/V Knorr and R/V Atlantis. When the sonar system is operational, multibeam swath data are collected for WHOI use only, unless previous arrangements have been made by the scientific party to purchase the data. Center-beam depth data, however, are regularly made available on-board the vessel as part of standard ship-system data provided during a cruise.

Multibeam data collected on WHOI vessels are proprietary to WHOI unless prior arrangement has been made for purchase of the data, in which case the data are available to the scientific party at sea. All multibeam swath data are archived in the WHOI MultiBeam Archive. Purchased data are proprietary to the PI for a period of two years and will not be distributed by WHOI. After two years, WHOI may distribute this data.

**Physical Oceanography - General comments**

Most data collected by the PO Department are digital in form and are controlled by the PI or a small, dedicated group such as the Float or SeaSoar groups. In several cases, these data are nicely organized and accessible via the web. Most data, however, are accessible only with prior knowledge of their existence (i.e. there are no metadata listing availability) and by directly contacting the PI. For older data, some migration to workstation platforms has occurred, but this has been a scattered effort.

All PO data types listed in the table have been assigned a priority level commensurate with category 2 -- "should be archived if resources are available". NSF stipulates that these data are to be archived with specific national data centers (see Appendix VI), although enforcement here has largely been nil in the past 5 years. The large programs (JGOFS, WOCE, etc.) are an exception in that they have fairly strict rules about submission, quality control, and access. Although WHOI, as an institution, need not be held responsible for locally archiving other PO data collected, WHOI investigators are technically in non-compliance. Furthermore, a deplorable result is increasing disorganization and scattering of the various datasets which have been collected through time. This situation negatively impacts on the ability of our scientists to do innovative research.

# APPENDIX VI

# EXISTING WHOI AND AGENCY DATA POLICIES

## APPENDIX VII

## NDSF ARCHIVING POLICY PROPOSED BY ad hoc SDAC

STATEMENT OF PURPOSE

The Woods Hole Oceanographic Institution (WHOI) maintains an archival system for oceanographic data, including geological samples and visual and digital information obtained using the vehicles and sensors of the National Deep Submergence Facility (NDSF). These vehicles currently include Alvin, Jason, Argo II, and the DSL-120 sonar. The federal funding agencies that support WHOI's Deep Submergence Operations Group (DSOG) provide funding to help support these archives. The funding agencies and the scientific community require that the data acquired using the NDSF facilities during scientific cruises be properly documented and archived for future scientific and educational use.

This archiving policy is intended to allow the WHOI DSOG to fulfill its obligations to funding agencies without unnecessarily compromising the Principal Investigator's (PI) right to sole use of the data for scientific purposes for a reasonable period of time, or compromising use of the data for public outreach and educational purposes in ensuing years.

Because there will be changes and improvements in imaging technology, recording media, and operational characteristics of the NDSF facilities in the future, details of this archiving policy may be revised from time to time. WHOI will provide the funding agencies and the Deep Submergence Science Committee (DESSC) with annual updates on the status of WHOI-archived NDSF data, together with any proposals for changes to improve data archiving.

PROTOCOLS FOR DATA HANDLING AND ARCHIVING

Each cruise that uses NDSF facilities has a Chief Scientist. This individual is designated by agreement between the Principal Investigator(s) and WHOI at least three months prior to the cruise. The Chief Scientist acts on behalf of the Principal Investigator(s) to provide a Cruise Prospectus to WHOI prior to the cruise, to conduct the expedition, to receive data from the DSOG and distribute it to the Principal Investigator(s) during the cruise, and to provide a Cruise Report to WHOI at the end of the expedition.

A copy of this Archiving Policy will be provided to the Chief Scientist at the time that he/she is designated. The Chief Scientist has the responsibility to assure that the entire shipboard scientific party is aware of this policy. Any questions should be directed to the WHOI DSOG Science Liaison.

Data are made available to the Chief Scientist during the cruise, with the exception of 35 mm transparency films which are provided after they are processed at WHOI. Data are available to other scientists at sea as directed by the Chief Scientist. Duplicating costs for additional copies of data will be charged to the individual receiving the data.

To assist in the proper archiving of data, and to assure that the data are useful to the scientific community, Chief Scientists are required to provide a complete Cruise

Report to WHOI Marine Operations at the end of the cruise. This report will outline events and data acquisition during the cruise such that archived data are fully documented and usable by any interested party, once they become accessible following the proprietary period noted below. The Cruise Report will be archived at WHOI.

Standard data acquired by NDSF facilities will be archived at WHOI as described below. Details about data acquisition (e.g., sources, processing, formats, recording media) are provided in the NDSF Users Manual.

## STANDARD DATA ACQUISITION BY NDSF VEHICLES

### Video Imagery

On each dive or lowering of Alvin, Jason, and Argo II, two video channels are recorded, one on each of two master recorders. All video data are duplicated at sea and first-generation copies are provided to the Chief Scientist for scientific research. The original video tapes are archived at WHOI.

### 35 mm Still Imagery

Standard operations provide for still photography using a 35mm camera mounted on Alvin, Jason, and Argo II. First-generation duplicates of all transparencies are provided to the Chief Scientist for scientific research, and the originals are archived at WHOI.

### Electronic Digital Still Camera Imagery

Digital images are collected using electronic still cameras on Jason and Argo II. With advance arrangements, they may also be collected on Alvin. These images are processed and recorded as standard image format files. The original media are archived at WHOI, and a copy is provided to the Chief Scientist for scientific research.

### Digital Sonar Data from the DSL-120 Sonar Vehicle

Data acquired using the DSL-120 sonar vehicle consist of side-looking sonar imagery and phase-difference bathymetric data. These data are processed at sea by DSOG personnel and the processed data are delivered to the Chief Scientist. The raw data tapes and a copy of the processed data files are archived at WHOI.

### Vehicle Navigation and Attitude Data, and Summaries

Vehicle navigation and attitude data are provided to the Chief Scientist as an ASCII string with all data columns identified. A tabular summary of all vehicle lowerings (geographic location, dates/times, etc.) is also provided as a text file and a hard copy. Copies of these data are archived at WHOI.

### Geological, Biological, and Geochemical Samples

DSOG personnel will document all samples (type, dive number, location, water depth, etc.) and their disposition. A copy of this documentation will be archived at WHOI together with any auxiliary data (e.g., descriptions, shipboard analyses) that are provided by the scientific party. Unless otherwise directed by the Chief Scientist, geological samples will be archived in the Seafloor Samples Laboratory at WHOI.

## DATA ACQUIRED USING THIRD PARTY TOOLS

Scientists may on occasion use sensors other than those in the normal DSOG inventory to accomplish their scientific objectives. Data acquired with these tools should be fully documented by the scientific party. The Chief Scientist will provide a copy of this documentation, together with a list of how the data are being distributed, to the DSOG for archiving at WHOI. The data are the property of the Principal Investigator who acquired them, and they will not be archived at WHOI except by special arrangement.

## ACCESS TO AND USE OF ARCHIVED DATA

During a proprietary period following the cruise, archived data may be accessed only by the Chief Scientist, or by others designated by the Chief Scientist in writing. The length of this proprietary period is the same as that dictated by the policies of the agency funding the expedition (e.g., normally 2 years for NSF and ONR funding). At the end of the proprietary period, all data are available for general use by scientists and educators, and for public outreach. Data will be made available to non-commercial users for the cost of reproduction and distribution.

Archived data are the property WHOI, and rights for commercial use of the data are vested in WHOI. Exceptions may be negotiated on a case-by-case basis if warranted. Such negotiations are to be conducted well in advance of the cruise on which the data will be collected. Otherwise the above policy will apply.

Principal investigators may use their discretion in distributing archivable images and data directly to the scientific and educational communities and to the news media for non-commercial use, or they may refer outside requests for such material to WHOI. PIs should insist that appropriate funding agency and institution acknowledgments accompany all such distributed material. (A standard credit might be: Courtesy of Woods Hole Oceanographic Institution and NSF; John Doe, Principal Investigator, XYZ University.)

## APPENDIX VIII

## CURRENT STATUS AND PROPOSED PROCEDURES FOR DATA HANDLING THROUGH MARINE OPERATIONS

### I. BACKGROUND

Any coordinated effort to enhance our archiving of oceanographic data should necessarily involve the individuals directly responsible for planning, supporting and executing voyages aboard our research vessels. The principal investigators, coordinators, schedulers and sea-going technicians are in the best position to define what data will be collected and subsequently to ensure that data is archived for future scientific use. In recent years, cruises aboard WHOI vessels have become more complicated and they increasingly involve multi-disciplinary projects and personnel with a wide variety of interests. As the complexity of scientific investigations grows, the task of organizing, coordinating, and disseminating details about a particular voyage becomes daunting. Planning for any particular aspect of the overall cruise program, such as data archiving, is therefore dependent on a viable communications network that allows all concerned parties access to timely information. Without such a structure, any archiving activity is liable to be incompletely planned and poorly executed.

### II. CURRENT STATUS AND PROPOSED MODIFICATIONS

WHOI Marine Operations is in the process of codifying a new communications framework to facilitate information exchange among all parties involved in data collection at sea. The following summarizes current procedures, with proposed modifications indicated in *italics*.

#### A. Pre-Proposal Stage

1. Principal investigators access the WHOI Marine Operations web site for technical information on vehicles, sensors, capabilities, and logistics. It is the intention of Marine Operations to make the use of web-based forms and databases as simple as possible to facilitate user compliance, cruise planning, and *an accurate accounting of data collected. In the future, Institution data archiving policy and requirements will be an integral part of information conveyed to prospective scientists on the web.*
2. PIs contact the WHOI science liaison with specific questions on how best to accomplish science programs with the vehicles. *Questions regarding the archiving policy can be addressed during this process.*
3. WHOI sends out notices to potential users several times a year, informing them of any important changes in web-posted information and reminding them about resources available on the web.
4. PIs fill out a web-based Cruise Planning Questionnaire, providing information pertaining to their proposed work. *Details about types of data to be collected will be requested in this form.*

#### B. Pre-Cruise Stage:

1. Following funding and scheduling of a program, a Cruise Synopsis is created from the original questionnaire requirements and posted on the web. *A special section will be*

*devoted to expected data collection and archiving concerns to avoid misunderstandings later in the process.* The PI provides updated information and any special logistical and technical requirements that could impact scheduling, staffing or technical preparations for the cruise.

2. A pre-cruise meeting is held with the PIs to exchange specific information regarding navigation requirements, dive/lowering plans, technical and equipment needs, and to identify scientific personnel for the cruise.

3. WHOI compiles the resulting information and relays it to shipboard technical personnel *and the WHOI Data Library.*

**C. Post-Cruise Stage:**

1. *Originals of specified data are sent to Scientific Archives; PI receives duplicates of this data. Metadata for materials not archived at WHOI will be logged. PI sends WHOI a complete Cruise Report for inclusion in the Scientific Archives.*

2. *After 2 years, cruise data in Scientific Archives is available for general scientific use.*

**D. Marine Technicians**

Seagoing marine technicians at WHOI are part of the Shipboard Scientific Services Group (SSSG), reporting to the Manager of Operational Scientific Services. They are responsible for maintaining and assisting in the operation of shared-use research equipment provided in support of cruise-specific scientific programs. These technicians also serve as a liaison between the science party and ship crew, a function which is particularly valuable to non-WHOI Chief Scientists. *It is proposed that the SSSG tech on WHOI vessels have the responsibility for implementing the Data Policy at sea and for coordinating efforts with Marine Operations and the Scientific Archivist on a routine basis. Tasks required of the technician will include maintaining a shipboard web site on a central server containing archiving policy documents, procedures and forms; providing scientific logs and software for tracking data collected; copying data such as video tapes and CD ROMs as appropriate; signing data out to PIs who intend to take original records off the ship; and shipping data and logs to the Scientific Archives. When possible, the shipboard technician and Chief Scientist will identify a member of the science party to work with the technician to catalog data collected. These duties do not represent a significant increase in work for the marine technician; on many cruises these functions are already part of the daily routine. Special SSSG tech focus on tracking and preserving data during the cruise, however, should help maintain a greater percentage of information for future use.*

## APPENDIX IX

## ACRONYM GLOSSARY

| | |
|---|---|
| ADCP | Acoustic Doppler Current Profiler |
| BIOMAPPER Environmental | Bio-Optical Multifrequency Acoustical and Physical Recorder |
| BPI | Bytes per inch |
| BU | Boston University |
| BUMP | Boston University Marine Program |
| CD | Compact Disk |
| CD ROM | Compact Disk, Read Only Memory |
| CIS | Computer and Information Services |
| CTD | Conductivity, temperature, depth |
| DAC | Data Assembly Centre (WOCE program) |
| DAT | Digital Acoustic Telemetry |
| DDL | Digital Data Library (Department of Geology and Geophysics) |
| DESSC | Deep Submergence Science Committee |
| DIU | Data Information Unit (WOCE program) |
| DLT | Digital Linear Tape |
| DSOG | Deep Submergence Operations Group |
| DSV | Deep Submergence Vehicle |
| GLOBEC | Global Ecosystems Dynamics (program) |
| GPS | Global Positioning System |
| HRP | High Resolution Profiler |
| IMETS | Improved Meteorological Packages |
| ISA | Independent Study Award |
| JGOFS | Joint Global Ocean Flux Study |
| LADCP | Lowered Acoustic Doppler Current Profiler |
| MBL | Marine Biological Laboratory |
| MBT | Mechanical Bathythermograph |
| MIT | Massachusetts Institute of Technology |
| MOCNESS | Multiple Opening and Closing Sampler Nets and Environmental Sampling System |
| MUDDIE | Metadata software for geological samples in the Seafloor Samples Laboratory |
| NGDC | National Geophysical Data Center |
| NMFS | Nation Marine Fisheries Service |
| NODC | National Oceanographic Data Center |
| NOSAMS | National Ocean Science Accelerator Mass Spectrometer (facility) |
| NDSF | National Deep Submergence Facility |
| NSF | National Science Foundation |
| ONR | Office of Naval Research |
| PDA | Personal Digital Assistant |
| PGR | Precision Graphic Recorder |
| PI | Principal Investigator |

| | |
|---|---|
| ROV | Remotely Operated Vehicle |
| SAC | Special Analysis Centres (WOCE program) |
| SDAC | Scientific Data Advisory Committee |
| SEA | Sea Education Association |
| SEDCORE | Metadata software for geological samples in the Seafloor Samples Laboratory |
| SQL | Structured Query Language |
| SSSG | Shipboard Scientific Services Group |
| STSA | Senior Technical Staff Award |
| UNIX | Computer platform |
| UNOLS | University-National Oceanographic Laboratory System |
| URN | Universal Resource Name (a permanent WWW address) |
| USGS | United States Geological Survey |
| VAX | Computer platform |
| WDC-A | World Data Centre |
| WHOI | Woods Hole Oceanographic Institution |
| WOCE | World Ocean Circulation Experiment |
| WWW | World Wide Web |
| XBT | Expendable Bathythermograph |

# DRAFT REPORT

# OF THE

# SCIENTIFIC DATA ADVISORY COMMITTEE

## 22 JANUARY 1998

**Committee Members:**

Brian Tucholke, Chairman (Department of Geology and Geophysics)
Richard Chandler (Marine Operations)
Ruth Curry (Department of Physical Oceanography)
David Glover (Department of Marine Chemistry and Geochemistry)
Robert Groman (Department of Biology)
Jonathan Howland (Department of Applied Ocean Physics and Engineering)
Colleen Hurter (MBL/WHOI Library)
Andrew Maffei (Computer and Information Services)

**Ex Officio:**

Cathy Norton (Director, MBL/WHOI Library)

**MBL/WHOI Library Committee Liasons:**

Daniel Fornari (Department of Geology and Geophysics)
Mark Kurz (Department of Marine Chemistry and Geochemistry)

7 January 1998
To Members of the SDAC

Here is a rough draft of the SDAC report. There are some internal inconsistencies that I've not had time to clear up, but I want to get this into your hands without further delay. The order departs from the original outline by design. Your comments on both organization and completeness of content will be welcomed.

Please review the entire document before our meeting on Monday. I've indicated in **bold** type the areas that need further discussion or specific input from committee members. Please email me your input for the areas requested and for any other significant additions, and bring a hard copy to the meeting on Monday. You can give me minor corrections in handwritten form.

I've indicated a few corrections that committee members need to make on tables in the appendices. Please bring corrected copies to the meeting. Thanks, and I'll see you on Monday.

Brian

COPY OF ORIGINAL

## APPENDIX VII - NDSF ARCHIVING POLICY PROPOSED BY ad hoc SDAC

**(1- I would like to include a brief summary of the main differences of opinion between the policy expressed here and the policy proposed by Barrie Walden. Dan, can you take a crack at that?**
**2- This policy needs to be updated to include geological and biological samples. We can discuss, Brian will write)**

### STATEMENT OF PURPOSE

The Woods Hole Oceanographic Institution (WHOI) maintains an archival system for oceanographic data, including visual and digital information obtained using the vehicles and sensors of the National Deep Submergence Facility (NDSF). These vehicles currently include Alvin, Jason, Argo II, and the DSL-120 sonar. The federal funding agencies that support WHOI's Deep Submergence Operations Group ( DSOG) provide funding to help support these archives through a portion of the overhead charge for operating the facility. The funding agencies and the scientific community require that the data acquired using the NDSF facilities during scientific cruises be properly archived for future scientific and educational use.

This archiving policy is intended to allow the WHOI DSOG to fulfill their obligations to funding agencies without unnecessarily compromising the Principal Investigator's (PI) right to sole use of the data for scientific purposes for a reasonable period of time, or compromising use of the data for public outreach and educational purposes in ensuing years.

Because changes and improvements in imaging technology, recording media, and operational characteristics of the NDSF facilities will take place in the future, the criteria and definitions set forth in this archiving policy may need to be revised from time to time. WHOI will provide the funding agencies and the Deep Submergence Science Committee (DESSC) with annual updates on the status of WHOI-archived NDSF vehicle data, together with any proposals for changes to improve data archiving.

### STANDARD DATA ACQUISITION BY NDSF VEHICLES

Each cruise that uses NDSF facilities has a Chief Scientist. This individual is designated by agreement between the Principal Investigator(s) and WHOI at least three months prior to the cruise. At the same time, the Chief Scientist will consult with all Principal Investigators to devise a list of parties to whom cruise data will be distributed and will provide this list to the DSOG. The Chief Scientist acts on behalf of the Principal Investigator(s) in conducting the expedition, receiving data from the DSOG and distributing it to the Principal Investigator(s) during the cruise, and providing required pre- and post-cruise reports to WHOI. The following describes the standard data a Chief Scientist can expect to receive from the various sensors on each vehicle, and what will be archived by WHOI.

Video Imagery

On each dive or lowering of Alvin, Jason, and Argo II, two video channels are recorded (currently in Hi8 S-video format), one on each of two master recorders. The inputs are from two of a variety of available imaging sensors that are operated and maintained in the DSOG inventory (currently one 3-chip CCD [charge couple device] camera, and various single-chip CCD cameras). The choice of the two inputs is at the discretion of the Chief Scientist. For each Alvin dive (nominal 6-hr. bottom time), six video tapes are generated: three 2-hr. tapes on each master recorder. During Argo II and Jason lowerings, video tapes are continuously recorded on each of the two master recorders. All video data are duplicated at sea and first-generation copies are provided to the Chief Scientist for scientific research. The original video tapes are archived at WHOI.

35 mm Still Imagery

Standard operations provide for up to two 100' rolls of 35mm transparency film per Alvin Dive, one roll per Jason lowering, and up to two rolls per Argo II lowering. These films are not developed at sea, with the exception of test strips that are processed to assure proper operation of the cameras. Developing is done at WHOI following the cruise (normally within 1-2 months) and duplicate transparencies are made. Costs for film, processing, and duplication beyond the standard operations noted above will be borne by the scientific party. First-generation duplicates of all transparencies are provided to the Chief Scientist for scientific research, and the originals are archived at WHOI.

Electronic Digital Still Camera Imagery

Digital images are collected using electronic still cameras on Jason and Argo II. With advance arrangements, they may also be collected on Alvin. These images are processed and histogram equalized at sea and are recorded as standard image format files on magnetic media. The original media are archived at WHOI, and a copy is provided to the Chief Scientist for scientific research.

Digital Sonar Data from the DSL-120 Sonar Vehicle

Data acquired using the DSL-120 sonar vehicle consist of side-looking sonar imagery and phase-difference bathymetric data. These data are processed at sea by DSOG personnel. The processed data are delivered to the Chief Scientist on Exabyte tapes in well documented, gridded formats with geographic referencing in UTM or latitude/longitude coordinates. Processing includes slant-range correction, speed correction, and navigational correction based on shipboard smoothed navigation. The raw data tapes and a copy of the processed data files are archived at WHOI.

Vehicle Navigation and Attitude Data, and Summaries

Vehicle navigation and attitude data are provided to the Chief Scientist as an ASCII string with all data columns identified. The data are provided on CD ROM or magnetic media, depending on the vehicle system and cost effectiveness in duplication and archiving. A copy of these data is archived at WHOI.

A tabular summary of all vehicle lowerings (geographic location, dates/times, etc.) will be provided to the Chief Scientist as a text file and a hard copy. Printed maps of vehicle navigation are also provided at appropriate scales as required by the Chief

Scientist during the cruise. A copy of the vehicle lowering summary will be archived at WHOI.

## THIRD PARTY TOOLS

Scientists may require that sensors other than those in the normal DSOG inventory be used on NDSF vehicles to accomplish their scientific objectives. These sensors fall under the definition of Third Party Tools, as described in separate guidelines provided by the federal funding agencies and DESSC. A scientist desiring to use a Third Party Tool on a NDSF vehicle must plan for this use well in advance (i.e., at least three months prior to sailing). As part of the planning process, the scientist must provide to the DSOG the requisite technical information needed to operate the sensor safely and effectively from the vehicle. The scientist is also responsible for providing recording media, for maintenance and repair of the equipment at sea, and for insurance in the event of equipment loss. The DSOG will assist in all ways possible both before and during the cruise to ensure the successful collection of data from Third Party Tools.

The Chief Scientist is responsible for providing to the DSOG Expedition Leader at the end of the cruise a copy of data acquired using Third Party Tools. This copy will be archived at WHOI as outlined in the following section.

## ARCHIVING PROTOCOLS

A copy of this Archiving Policy will be provided to the Chief Scientist at the time that he/she is designated (i.e., at least three months pre-cruise). The Chief Scientist has the responsibility of assuring that the entire shipboard scientific party is aware of this policy. Any questions should be directed to the WHOI DSOG Science Liaison.

With the exception of 35 mm transparency films, data are made available to the Chief Scientist during the cruise. The 35 mm transparencies are processed at WHOI, normally within 1-2 months following the end of the cruise, and first-generation duplicates are sent to the Chief Scientist as soon as they are available. Data are also available to other scientists, as directed by the Chief Scientist. Duplicating costs for additional copies of data will be charged to the individual receiving the data.

During a proprietary period following the cruise, archived data may be accessed only by the Chief Scientist and his/her designees. The length of this proprietary period is the same as that dictated by the policies of the agency funding the expedition (e.g., normally 2 years for NSF and ONR funding). At the end of the proprietary period, all data are available for general use by scientists and educators, and for public outreach.

## CRUISE REPORTS

To assist in the proper archiving of data, and to assure that these data are useful to the scientific community, Chief Scientists on all cruises that use NDSF vehicles are required to provide a complete cruise report to WHOI Marine Operations within two (2) months post-cruise. This report will outline events and data acquisition during the cruise

such that archived data are fully documented and usable by any interested party. The report will be archived at WHOI.

## ACCESS TO DEEP SUBMERGENCE DATA

### Access by the Scientific and Educational Communities and News Media

WHOI-archived images and data collected by NDSF vehicles represent a valuable resource for scientific research and public education. As a not-for-profit institution and a world leader in the collection of undersea data and imagery, WHOI has a responsibility to properly archive data and make them widely available. Archival methods and policies allow this resource to be indexed and preserved and to be readily accessible for multiple applications. Extensive use of archived material by scientists, educators and news media is encouraged and will be facilitated by WHOI archivists, with the goal of enhancing knowledge and popular appreciation of the oceans, scientific research, and deep submergence technology. Data that are not proprietary will be made available to non-commercial users for the cost of reproduction and distribution.

### Commercial Rights in Data Acquired with Public Funding

Commercial rights for archived material that was obtained by DSOG-maintained sensors through funding by public agencies are retained and administered by WHOI on behalf of the agencies that support the NDSF. Commercial rights for archived material that was obtained by Third Party Tools funded by public funds will be shared equally by WHOI and the investigator providing the Third Party Tool.

Commercial entities are expected to provide funds beyond the cost of duplication and distribution in accordance with expected profits. WHOI (together with the investigator providing a Third Party Tool, when appropriate) will negotiate these fees on a case-by-case basis. All proceeds accruing to WHOI from commercial use of imagery or other deep-submergence data will be used to maintain the WHOI archives and to improve the scientific capabilities of the NDSF vehicles.

### Commercial Rights in Data Collected by Third Party Tools Funded by Private Funds

Commercial rights for archived material that was obtained by Third Party Tools funded by a private party, but collected on an expedition otherwise funded by public funds, will be shared equally by WHOI and the private party. WHOI, together with the private party, will negotiate these fees on a case-by-case basis. All proceeds accruing to WHOI from commercial use of these data will be used to maintain the WHOI archives and to improve the scientific capabilities of the NDSF vehicles.

### Commercial Rights in Data from Expeditions Funded Exclusively by Private Funds

When expeditions are funded exclusively by private funds, commercial rights to images and data collected with Third Party Tools will be vested in the party providing those tools. As part of the terms for use of the NDSF vehicles, one copy of the

images/data will be archived in the Scientific Archives for scientific and educational, non-commercial use only. These images/data will be proprietary to the party providing the Third Party Tool(s) for a period of time negotiated prior to the cruise (normally two years). Commercial rights to images and data collected from cameras and sensors maintained routinely by the DSOG as assets of the National Deep Submergence Facility vehicles are retained and managed by WHOI. These images and data are proprietary to the party funding the expedition for a period of two years following the end of the cruise.

## EXCEPTIONS

Exceptions to the above policy may be negotiated on a case-by-case basis when warranted. These negotiations are to be conducted well in advance of the cruise on which the data will be collected. Otherwise the above policy will apply.

## RIGHTS OF PRINCIPAL INVESTIGATORS

Principal investigators may use their discretion in distributing archivable images and data directly to the scientific and educational communities and to the news media for non-commercial use, or they may refer outside requests for such material to WHOI. PIs should insist that appropriate funding agency and institution acknowledgments accompany all such distributed material. (A standard credit might be: Courtesy of Woods Hole Oceanographic Institution and NSF; John Doe, Principal Investigator, XYZ University.)