USGS
science for a changing world

Collections Management Plan for the U.S. Geological Survey Woods Hole Coastal and Marine Science Center Data Library

Open-File Report 2015–1141

U.S. Department of the Interior
U.S. Geological Survey

USGS
science for a changing world

Collections Management Plan for the U.S. Geological Survey Woods Hole Coastal and Marine Science Center Data Library

By Kelleen M. List, Brian J. Buczkowski, Linda P. McCarthy, and Alice M. Orton

Open-File Report 2015–1141

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2015

Suggested citation:
List, K.M., Buczkowski, B.J., McCarthy, L.P., and Orton, A.M., 2015, Collections management plan for the U.S. Geological Survey Woods Hole Coastal and Marine Science Center Data Library: U.S. Geological Survey Open-File Report 2015–1141, 16 p., http://dx.doi.org/10.3133/ofr20151141.

ISSN 2331-1258 (online)

Acknowledgments

The authors would like to thank Nancy Soderberg for her wealth of advice and experience in the creation and operation of the Woods Hole Coastal and Marine Science Center's Data Library. We would also like to thank Fran Lightsom for her guidance in the production of this report. This report has benefited from critical reviews by Carolyn Degnan of the U.S. Geological Survey and David Sherman of the Woods Hole Oceanographic Institution.

Contents

Figures

Tables

Abbreviations

CD	compact disc
CDP	common depth point
DVD	digital versatile disc
FAN	field activity number
ID	identification number
MOF	Marine Operations Facility
NARA	National Archives and Records Administration
PI	principal investigator
USGS	U.S. Geological Survey
WHCMSC	Woods Hole Coastal and Marine Science Center

Collections Management Plan for the U.S. Geological Survey Woods Hole Coastal and Marine Science Center Data Library

By Kelleen M. List,[1] Brian J. Buczkowski,[1] Linda P. McCarthy,[1] and Alice M. Orton[2]

Abstract

The U.S. Geological Survey Woods Hole Coastal and Marine Science Center has created a Data Library to organize, preserve, and make available the field, laboratory, and modeling data collected and processed by Woods Hole Coastal and Marine Science Center staff. This Data Library supports current research efforts by providing unique, historic datasets with accompanying metadata. The Woods Hole Coastal and Marine Science Center's Data Library has custody of historic data and records that are still useful for research, and assists with preservation and distribution of marine science records and data in the course of scientific investigation and experimentation by researchers and staff at the science center.

The data accession and retention policies employed by the Woods Hole Coastal and Marine Science Center Data Library are based on scientific need and the National Archives and Records Administration standards for Federal records retention. Criteria for inclusion of data and records into the Data Library, the scope of the Data Library holdings, and operating procedures for the management and running of the library are designed to support the research operations of the U.S. Geological Survey.

This report explains the roles and detailed responsibilities of library and scientific staff, and provides step-by-step instructions for managing the collections of the Woods Hole Coastal and Marine Science Center Data Library.

Introduction to the Woods Hole Coastal and Marine Science Center Data Library

The U.S. Geological Survey (USGS) Woods Hole Coastal and Marine Science Center (WHCMSC) is colocated with the Woods Hole Oceanographic Institution on the Quissett Campus in the village of Woods Hole, Massachusetts. The WHCMSC maintains laboratory facilities and a samples repository on this campus, as well as a warehouse facility, the USGS Marine Operations Facility (MOF), approximately 2 miles north of the Quissett Campus in the town of Falmouth, Mass. The MOF warehouse encompasses storage and equipment staging areas and the WHCMSC **Data Library**.[3]

Mission statement.—The mission of the WHCMSC Data Library is to organize, preserve, and make available for research purposes a historical record and archive of scientific **data** collected by WHCMSC personnel, as well as institutional **records** used by researchers. The Data Library provides access to

[1]U.S. Geological Survey
[2]Geocon, Inc.
[3]Terms in bold are defined in the glossary.

scientific datasets and record series for reuse in future research by USGS scientists, the public, industry, policy makers, outside researchers, and academic institutions.

The WHCMSC Data Library provides a local repository for original data and related support documentation. Long-term preservation of these records enables further scientific investigation based on previous research. Collecting marine scientific data is expensive and often unrepeatable; thus it is imperative that the data are preserved in a reusable form for use in future investigations. The Data Library holdings support the institutional memory of the WHCMSC and support the efforts of future researchers and historians.

Scope of the Collections in the Data Library

The WHCMSC Data Library includes a library of resource materials and published records, in addition to a facility dedicated to preserving and providing scientific data. The Data Library maintains a collection of USGS publications, including professional papers, bulletins, circulars, open-file reports, and various topographic, bathymetric, and geologic map series. It also provides a small reference collection of outside publications and materials for use in science planning. This collection includes National Ocean Service nautical charts and bathymetric maps, professional association publications, and educational outreach materials (see "Library Organization" section for more information about these collections and how they are categorized). The Data Library also currently provides select datasets online. Many digital collections of **field activity** data are available to WHCMSC researchers through USGS internal Web pages, and datasets approved for public release are available through USGS public-facing Web pages.

Archiving data.—The WHCMSC Data Library has a responsibility to preserve and provide several categories of data. These categories represent different stages in the lifecycle of data or records, which begins with data acquisition or generation and results in published data and analytical results.

Final processed data are datasets derived from original **raw data** and have reached a level of completion such that researchers and authors have determined they are ready for publication and distribution. These final processed data are considered the end member of a suite of processed versions derived from the raw data. The WHCMSC Data Library does not retain interim versions that result from various processing steps, unless they qualify as **essential support data**.

Essential support data are neither raw nor final processed but are determined to be invaluable as a resource for understanding the nature of the data and aiding in their interpretation and further use. These data may consist of interim processing steps or of ancillary data valuable to research, public awareness, or even litigation. In the case of litigation, every processing step and note is retained for future reference. Essential support data often provide the only record of data and results available to the Data Library until raw and final processed data are submitted for retention. Decisions regarding datasets and their classification as essential support data can come from WHCMSC researchers and management, the USGS, the U.S. Department of the Interior, or higher levels within the Federal Government.

Geology Discipline Research Records Schedule

As part of the USGS's responsibility to the public, the Federal research records which are created in the course of conducting business are to be strictly preserved and managed in accordance with the Federal Records Act (44 U.S.C. §§3101–3102). The WHCMSC Data Library is tasked with the management, preservation, and **disposition** of these records in accordance with the Geology Discipline

Research Records Schedule of U.S. Geological Survey Manual 432–1–S5 (U.S. Geological Survey, 2009).

This records schedule defines three categories of research records:

1. significant research records (permanent)
2. secondary research records
3. minor research records

Significant research records are defined as "permanent" records according to the Geology Discipline Research Records Schedule. Permanent records are characterized as "datasets that are irreplaceable, relevant to the USGS mission, and in a condition which allows future use" (U.S. Geological Survey, 2009). These records include collections of related data derived from field activities, laboratory analyses, computer models and simulations, including data donated by outside organizations, and accompanying **metadata**. The majority of the Data Library holdings consist of permanent records. Because a study area can be revisited decades later, the retention of these data and records has proven to be invaluable both for planning further scientific investigations and also as a baseline for studying changes through time. The Geology Discipline Research Records Schedule requires that permanent significant research records be transferred to the National Archives and Records Administration (NARA) when the USGS "has no further use for them" (U.S. Geological Survey, 2009).

The continued value of these data for current and future research is the reason for the WHCMSC Data Library. Seismic profiles and other raw data serve as a historical record and often a starting place for new field work. Without the retention and accessibility of this information within a local repository, future research can be jeopardized because of the time and monetary constraints on collecting this data framework again.

Secondary research records focus on immediate, short-term tasks and are not associated with scientific investigations of larger scope. Secondary research records have implied value for future scientific investigations and should be reviewed before disposition in case they have been incorporated into later projects that warrant permanent retention. These records are sometimes gray literature or scientists' papers that are **accessioned** into the Data Library's Reference Resources collection as background materials, descriptions of new techniques, or information about physical samples. Secondary research records might have limited long-term usefulness at the WHCMSC, and they should be evaluated for offsite storage at NARA after five years, consistent with the Geology Discipline Research Records Schedule.

Minor research records include records relating to narrowly focused, short-term tasks that address specific, local problems. These records are not part of investigations of larger scope and provide no potential for future investigation. These records are generally the responsibility of scientists and not accessioned by the Data Library. However, records such as rough drafts, publication copies, and policy statements may be accepted into the Data Library if they meet criteria for use as educational or reference materials.

Roles and Responsibilities

Data Library Staff

The data librarian is responsible for the operation, organization, and management of the WHCMSC Data Library. The data librarian reports to a data manager and supervises the day-to-day activities and assignments of Data Library staff and volunteers. The data librarian is tasked with the preservation of data, publications, and research records. In pursuit of these goals, the data librarian

evaluates incoming data for accession (see "Accession of Data and Materials" section), reviews collections for duplicate datasets, oversees the digitization of holdings, handles data requests, follows up on retrieval of loaned records, and **deaccessions** (see "Deaccession" section) records and collections as appropriate. The data librarian initiates, coordinates, and preserves video-recorded oral histories of key, long-term WHCMSC personnel for the preservation of institutional memory. The data librarian is a designated records management specialist and assists USGS researchers in taking care of research records to be consistent and compliant with the Geology Discipline Research Records Schedule.

Research Staff

Principal Investigator

The primary role of the **principal investigator** (PI) within the WHCMSC Data Library system is to provide the Data Library with contributions of Federal records and data of enduring value to the USGS. Upon completion of a field activity, laboratory analysis, or modeling run the PI or designated representative is responsible for providing an unaltered, unprocessed (raw) copy of data collected or generated to the Data Library for immediate **inventory** and accession into the library holdings. Upon completion of a research project, the PI must provide all related scientific papers, notes, printouts, and digital records that explain data acquisition, processing, analysis, or other elements essential to understanding any final reports or published products. The PI must make an appointment with the data librarian to discuss accession of scientific data into the Data Library. The PI provides the **field activity number** (FAN), laboratory analysis number, modeling project number or donated data number; any other pertinent identifiers, such as an older cruise identification number (ID); the data (raw and final processed); and all relevant metadata. It is also the PI's responsibility to let the data librarian know when the data have been published and can be released to the public.

Retiring and Departing Scientists

Retiring and departing personnel are responsible for identifying Federal records in their possession and must follow the approved records management schedule to organize and prepare their records for archiving. See appendix 2 of the Geology Discipline Research Records Schedule for guidance in identifying Federal records (U.S. Geological Survey, 2009).

Retiring and departing employees need to meet with the data librarian to help plan for the accession of the scientific Federal records in their custody in accordance with their responsibilities as defined in the Geology Discipline Research Records Schedule. According to the schedule, the retiring or departing employee must provide a file plan to the records management specialist—which in this case is the WHCMSC data librarian. All permanent records must be so classified, and if they are research records they must be associated with a FAN, laboratory analysis number, modeling project number, or donated data number. All field activity-related records, laboratory notes, and logs are transferred into the custodianship of the Data Library.

Retiring and departing employees are also required to complete an **exit survey form**. The WHCMSC Data Library requires the identification of all field activities in which the employee participated and the identification and current location of all related data (files, images, hard drives). Each dataset and record should be accompanied by its corresponding FAN, laboratory analysis number, modeling project number, or donated data number; status; and a point of contact for followup inquiries and continuity of operations. A curriculum vitae with field activity annotations in an electronic form, preferably in a standard bibliographic format, should also be supplied to the data librarian.

Library Patrons

Records and resources in the Data Library are made available to WHCMSC personnel and to the general public upon request. When patrons require the retrieval of data stored in the Data Library, they are asked to email a request with specific information regarding the desired materials, especially FANs or cruise IDs, laboratory analyses, or published models, if known. Patrons make requests of the WHCMSC Data Library in search of historical data that support current or planned scientific research. If an item of interest is not available online, the data librarian finds the physical original, scans it, and supplies a digital and (or) paper copy to the requestor. Data requests from the public for data and results that have been published or otherwise publicly released are also fulfilled by the WHCMSC Data Library.

Data Library Operations

Library Organization

The WHCMSC Data Library maintains three organizational categories for its holdings: distribution resources, reference resources, and science center data. These categories are an aid to determining storage, access, and preservation needs for data and records in the Data Library.

Distribution resources are defined as non-Federal records and items maintained by the library specifically for public distribution purposes. These include educational and outreach materials such as maps, project fact sheets, and copies of WHCMSC-authored publications.

Reference resources can be Federal or non-Federal records and are maintained by the WHCMSC Data Library for historical and (or) research purposes. Non-Federal records include non-USGS-authored publications, maps, and data. Federal records classified as reference resources retained by the Data Library include administrative records of WHCMSC projects such as planning documents, pertinent emails, diagrams, and more. Reference resources categorized as Federal records must be reviewed according to the guidance provided by the USGS Geology Discipline Research Records Schedule (U.S. Geological Survey, 2009) and evaluated for continuing usefulness to the USGS.

Woods Hole Coastal and Marine Science Center data are records or datasets collected or created by, or in cooperation with, WHCMSC employees or by using WHCMSC equipment. These data can include, but are not limited to, notes and logs, photographs, video, navigation data, bathymetry, seismic data, sonar data, oceanographic measurements, resistivity data, and other geophysical and geochemical data. This section of the Data Library is consistent with the well-accepted practice of maintaining a research data repository that includes unique, institutionally derived holdings. Data donated by sources outside of the WHCMSC and accepted by the Data Library for preservation and distribution are also considered science center data.

Within the Data Library, science center data are accompanied by associated metadata in compliance with USGS Instructional Memorandum IM OSQI 2015–03 (U.S. Geological Survey, 2015b), which states, "Metadata must accompany all USGS scientific data, software, and other information products described in this policy that are approved for release. The content and format of metadata depend upon the type of data, software, or information product being described."

The WHCMSC Data Library accepts all types of data storage media and has a dedicated facility equipped to handle diverse organizational and storage requirements in a controlled environment to prolong the life and physical integrity of these media. Data may be in any file format and on any medium (paper; videotape; Mylar; or digital media, including compact disc [CD], digital versatile disc [DVD], hard drive, or tape). The Data Library is equipped with large rolling storage systems, map

cabinets, rolled media storage cabinets and boxes, tape drawers, photographic library boxes, and record-storage file boxes. All related project records and contributions are maintained in the WHCMSC Data Library and organized by using an associated FAN, laboratory analysis number, modeling project number, or donated data number.

Accession of Data and Materials

For the purposes of this report, accession is defined as the evaluation and acceptance of data and records into the WHCMSC Data Library. The following are the steps usually applied to materials considered for acceptance into the WHCMSC Data Library.

Step 1: Intake

The **intake** process begins with a record or records coming to the attention of the data librarian for possible accession into the WHCMSC Data Library. Frequently, the initial review of these records is conducted at the donor's office where the data are stored.

Through conversation and evaluation, the data librarian determines if the data have been collected by the WHCMSC or were obtained from another source.

Data collected by the WHCMSC must have a corresponding FAN, laboratory analysis number, or modeling project number. If only an alternate identification designation is available, the data librarian generates a FAN, laboratory analysis number, or modeling project number for the data.

If the data are donated from a source outside the USGS, their appropriateness for inclusion into the Data Library must be determined. Do they support the mission of the science center; and if so, are they authorized for use and redistribution? If not, can the donation be useful for outreach activities or as resource data? If the donation is not authorized for distribution and cannot be used for outreach, it cannot be accessioned into the library and is returned to the originator; disposition should be discussed if the donation cannot be returned to the originator.

Acceptable donated and science center data associated with FANs, laboratory analysis numbers, modeling project numbers, or donated data numbers must pass several criteria for appropriateness:

1. Are the data raw, final processed, or essential to the analysis process?
2. Are the data type and format identifiable and readable?
3. Are science center personnel willing to be responsible as **scientific authorities** for the data?

It is important to have a scientific authority as a point of contact to address any science questions that may be asked of the data librarian and to be consulted regarding the data's long-term usefulness to the USGS.

If the answer to any of the three previous questions is "no," disposition of the material should be considered. If the data meet the three conditions, they are **checked in** as a Data Library project and join the data records waiting for inventory evaluation.

At this phase, the data record is accepted only for intake and still must be evaluated for acceptance into the Data Library holdings. It is not guaranteed for future accession. It is in a temporary holding and sorting area organized by **project numbers**. This step has become unavoidable because of the volume of data, limited library space, and staffing needs. To keep the library from becoming overrun with data in need of evaluation, classification, and inventory accession, the data are put in a box and assigned a project number from a master spreadsheet. The FAN, laboratory analysis number, modeling project number, or donated data number, along with any other general metadata for the project content, are entered on the master spreadsheet. After the record is assigned a project number, it is considered checked in to the Data Library. This process is diagrammed in figure 1.

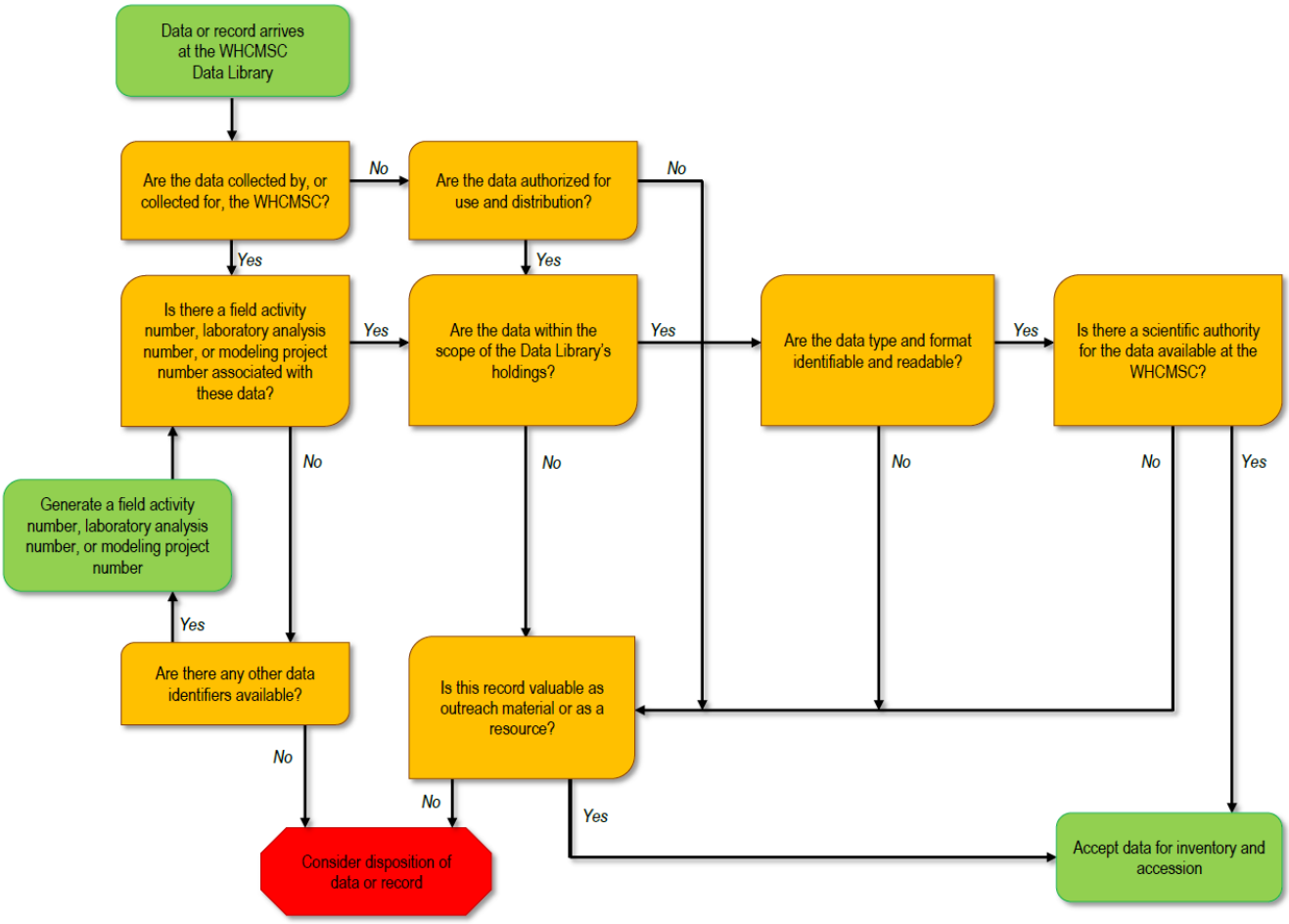

Figure 1. Decision tree showing the questions used to determine if records, data, and related materials are suitable to be accessioned into the Woods Hole Coastal and Marine Science Center's Data Library collections. This tree also presents options and actions required in relation to answers to these questions. WHCMSC, Woods Hole Coastal and Marine Science Center.

Step 2: Inventory and Evaluation

In this step, checked-in data from a project are brought forward for evaluation. If donated data checked in to the Data Library are proprietary to another organization or person, the data librarian or responsible center staff must secure permission for the document's release. These donated data then move into the **catalog** step. If document release permission cannot be obtained for donated data, the proprietary records are returned to the donating agency or individual.

Checked-in science center data then undergo an evaluation process to determine if the records will be accessioned into the holdings. The first step to check in data is to research all available records and data related to the FAN, laboratory analysis number, modeling project number, or donated data number identified during the intake process for these records, and determine if the data currently exist in the library holdings. If the field activity Web page or Data Library database indicate copies of the data are already in the Data Library, the newly arrived data must be evaluated against the current library holdings to determine if they add value to the collections. Are the new data in better condition? Is the

medium of the new data more current? Are the file formats the same? If the data are in a different format, does the format add value to the holdings? If analysis determines that the data add value, a decision must be made whether it is appropriate to keep both copies or to replace the existing data record with the newly acquired data record.

After it is determined that it is appropriate to store the record in the WHCMSC Data Library, the record is further evaluated for accession. Some items are appropriate for inventorying (**inventory numbers** are assigned), while other items are cataloged or directly shelved (fig. 2).

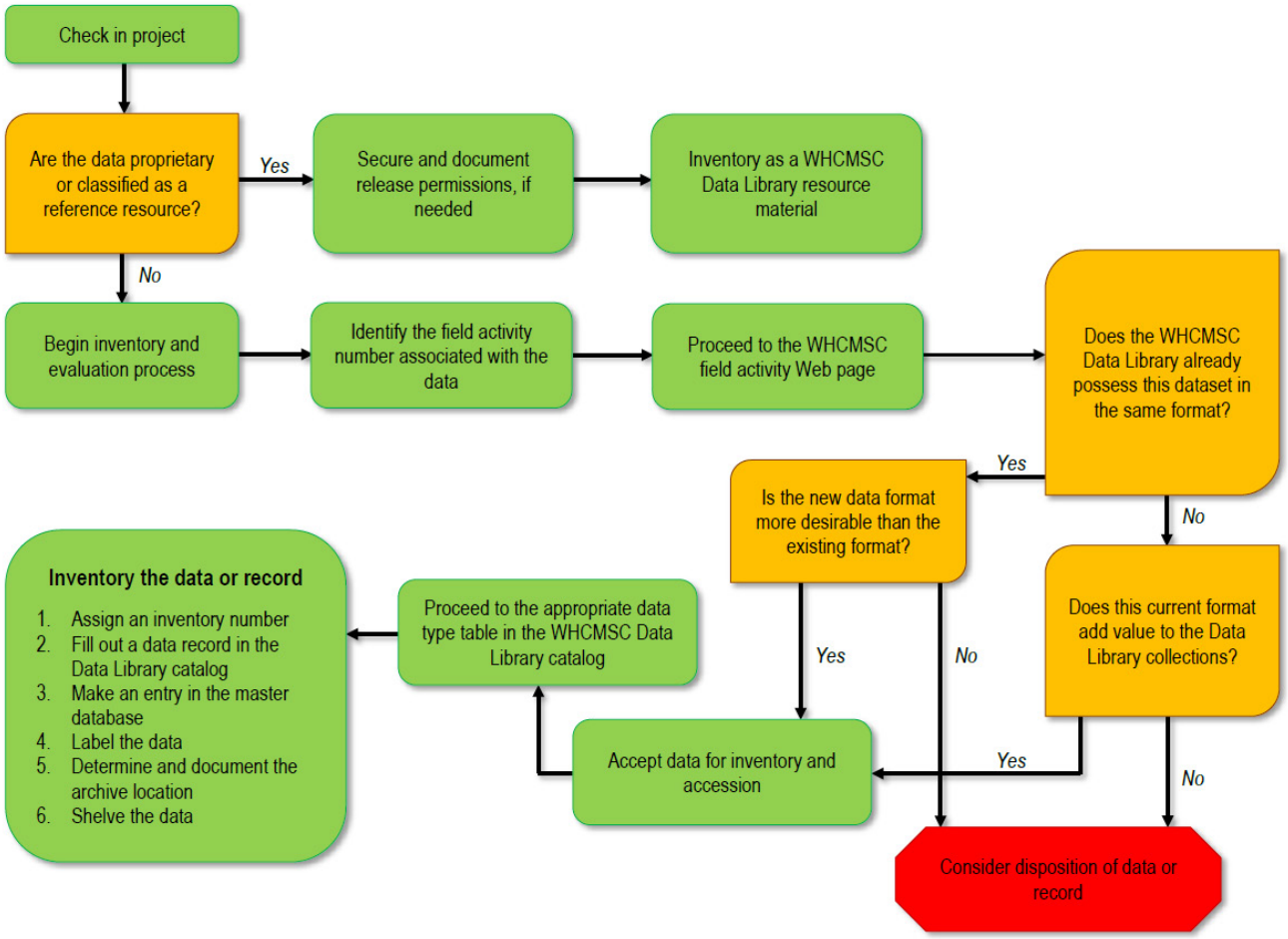

Figure 2. Flow chart detailing the inventory and evaluation procedures implemented by the Woods Hole Coastal and Marine Science Center's Data Library and applied to all records, data, and related materials deemed appropriate for accession. WHCMSC, Woods Hole Coastal and Marine Science Center.

Step 3: Catalog or Inventory

These steps occur when the record has been evaluated to be accepted into the WHCMSC Data Library. Items that are assigned inventory numbers are expected to eventually become publicly advertised and available. Other types of records and documents are cataloged and shelved, and access to them is assessed on an individual basis.

Catalog

The catalog process begins by identifying the type of data, gathering required format-specific metadata for the record, and making an entry into the appropriate database table (table 1). The first step is to identify what type of data the record represents. The categories of records to be cataloged include five types of data:

1. Educational materials such as science teacher toolkits, posters, and experiments may be used in outreach or provided to educators upon request.
2. Nautical charts are kept for distribution to the research staff and the general public.
3. Reference materials consist of pertinent texts used by WHCMSC staff as a reference library.
4. Background information pertaining to research vessels is identified and cataloged. Information about the vessel or platform often proves helpful in identifying cruise information such as IDs and timeframes, equipment deployed, and research data collected.
5. Topographic maps are cataloged and maintained in support of current or historical WHCMSC projects because of their value to researchers and the general public.

Table 1. Categories of cataloged items retained by the Woods Hole Coastal and Marine Science Center Data Library for use as references and educational materials.

Data Library database table name	Type of data
Education	Educational outreach resources
Nautical	Nautical charts
Reference	Reference resources
Ships	Background information about research vessels
Topmap	Topographic maps

Inventory

Inventoried records include all raw and final processed data collected by WHCMSC researchers, including logbooks, navigation data, track charts, seismic folders, common depth point (CDP) seismic records, CDP track charts, submarine video, and photographs (both analog and digital formats). To begin the inventory process, the data librarian gathers or creates metadata that will be entered into the Data Library database. An inventory worksheet is used to organize these metadata (fig. 3; the abbreviations are listed in table 2). Metadata are entered into the corresponding database table with each unique piece of information entered into its own record. As an example, if a collection of CDs is submitted to the Data Library and approved for retention, each CD is entered separately with its own metadata. The Master database table fields are then filled in, beginning with an inventory number. The WHCMSC Data Library uses several series of inventory numbers based on the types of records (table 3).

Accession of Data Trackchart TRKCHART Received by: LPM
From: Anne Trehu - PI Date: Date Record Created Nov. 14, 2014

TRKCHART		MASTER		Compass Record	1983-008-FA-LN-003
Gen Descript	Trackchart	Gen. Descript.	Trackchart	FAN	1983-008-FA
INVNUM	TK00821	Scan? Who:Date:Where		EQUIPMENT	Integrated Navigation System
DATA TYPE	NAV	INVNUM	TK00821	NAME	Trackchart
PLATFORM	Gyre	PLATFORM	Gyre	DESCRIPTION	Navigation track line plot for Lines 1, 1A, 2, 3, 4, 5
CRUISE	GY-83-12	FIELD_ACTIVITY	83008	CONTACT	Linda McCarthy
SERIAL	83008	CRUISE	GY-83-12	PHYSICAL LOC.	WHSC Data Library Storage area.
LOCATION	T-18	YEAR	1983	DIGITAL LOC.	
PROJ		CRUISE_ID	Gyre19830822	MEDIA	Mylar
SCALE		DATA_TYPE	NAV	FORMAT	Trackchart
FORM	Mylar	FORM	Trackchart	AMT/SIZE	1
SOURCE	Plot	AMT	1	DATA PROCESSOR	
MAPID	1	SOURCE	Computer	PUBLIC ACCESS	Yes
SET	1	NATURE	ORIG	DATA REVIEWED?	NO
AREA	Offshore - Fl/Cardina	FORMAT	Mylar	DataToBeProcessed?	
		LOC	T-18	FINAL REPOSITORY	WHSC Data Library
REMARKS	Lines 1, 1A, 2, 3, 4, 5	DETAIL	TRKCHART (Table)	URL	
		REMARKS	Lines 1, 1A, 2, 3, 4, 5	DATE PUBLISHED	
				VERSION	
				COMMENTS:	
NOTES		NOTES		Supporting Docs ?	Yes
				Supp.DocsPhys.Loc	WHSC Data Library
				Supp.Docs.Digital Loc	
				Metadata?/loc. /reviewed?	
				Navigation Info: Type Station Horizontal datum Vertical Datum Start-End Time GMT offset	
				LOCATION MAP	
				STATUS	
				FILES	

Figure 3. Example accession worksheet for a track chart. The sheet requires metadata that will be applied to three collections management tables administered by the data librarian: TRKCHART; MASTER; and Compass Record. Abbreviations are listed in table 2.

Table 2. Abbreviations relating to the Woods Hole Coastal and Marine Science Center's Data Library's accession worksheet.

[The worksheet is in figure 3]

Abbreviation	Definition
TRKCHRT	track chart
Gen Descript	general description
INVNUM	inventory number
SERIAL	serial number
PROJ	project number
MAPID	map identification number
FORM	format
AMT	amount
LOC/LOC./loc.	location
FAN	field activity number
URL	uniform resource locator
Docs	documents
Supp.DocsPhys.Loc	physical location of supporting documents
Supp.Docs.Digital Loc	digital location of supporting documents
Info	information
GMT	Greenwich Mean Time

Table 3. Explanation of inventory numbering system as defined in the INVNUM table in the Woods Hole Coastal and Marine Science Center's Data Library.

[INVNUM, inventory number; CDP, common depth point]

Type of record	Subtable	Inventory number prefix
CDP	CDPREC	CR
CDP track lines	CDPTRACK	CT
Film	FILM	(6-digit number—no prefix)
Maps	MAPS	MP
Optical disks	OPTDISK	CD
Seismic records	SEISRECS	FD
Submarine video	SUBVIDEO	(6-digit number—no prefix)
Seismic analog tapes	TAPSA	SA
Track charts	TRKCHART	TK
Seismic and sidescan data on video	VIDEOSA	VS

Data records that do not have subtable entries include navigation logs, seismic profiles, observer logs, photographs, and PI notes and records. Records in the Master database table contain general information regarding the data, whereas the subtables for each data type include more detailed information about the data. The Master database is used as the overarching index to all data stored and shelved in the Data Library.

Deaccession

Deaccession is the process used to formally remove holdings from the WHCMSC Data Library's custody and collections. This process closely adheres to the Geology Discipline Research Records Schedule. Deaccessioning materials is an evaluation of the future usefulness of the records (see "Deaccession Criteria" section) and requires appropriate documentation and permissions.

Data formats can reduce the value of records because of decomposition of the storage media or inability to access them. Value can also decrease if the materials become available in some other way, for instance if another agency makes them available digitally. These conditions may qualify records for deaccession by the Data Library.

Determining if records are appropriate for deaccession should proceed with careful consideration of other means of access to the materials, other storage formats, and other storage locations that might be available. An appropriate PI (or scientific authority) should be consulted about the importance of maintaining data and records in a particular format. Permissions and authorities of all advising scientists should be carefully documented for future reference.

An additional criterion is the continued use of the records by the USGS. Significant (permanent) research records should be maintained in a usable form until they are transferred to NARA upon determination that the records are no longer of use to the USGS. Annually, Federal records that are classified as nonpermanent secondary research records should be evaluated for further retention and appropriate disposition.

Deaccession Criteria

When records are considered for deaccession, it is important to assess the value of the collection for further research, educational, or outreach potential:

- Consider the condition of the media, accessibility of the media, and any previous data migrations (such as data migration from tape media to computer server, scanning of paper products to photographic images, and so forth).
- If materials are not data (general reference materials, publications, educational items, maps, or charts), are the records now available through another agency or in a more modern format? For instance, is a record previously released in book format now available digitally?
- Are the data and records accessible to local researchers from other sources, or is access to them restricted?
- If the data are proprietary, check with the donating institution, if possible, or with projects at the WHCMSC involved in closely related research to determine whether the records still need to be stored locally.

If these conditions are evaluated, and the data and records are determined to be of no further use to WHCMSC research, it is appropriate to deaccession the records.

Disposition

Disposition is the process of physically removing data and records that are deaccessioned or determined inappropriate for inclusion in the Data Library. For permanent Federal records, disposition means transfer to NARA. Other data and records can be donated or destroyed. When deaccessioned records are transferred, donated, or destroyed, the disposition should be noted in all relevant Data Library database tables.

If items to be deaccessioned are proprietary data, the records should be returned to the original donor, and the transfer should be documented. If the data were previously donated to the WHCMSC but the Data Library now has permanent custody of the records, they should be handled as if collected by WHCMSC. If data consist of permanent Federal records and have been evaluated as having no further usefulness to the USGS, these records qualify for permanent transfer to NARA, and steps should be taken to turn over these materials to the appropriate NARA archive.

In all cases of disposition, it is imperative to document the conditions and the reasons for the deaccession. All media migrations and new locations of data are documented. Permissions and authority given for the deaccession including names, any objections, person's relationship to the data, dates, and so forth are recorded and tracked.

Nonpermanent record collections no longer needed for current research or operations, not determined to be useful for further scientific investigations, and not accepted by NARA must be disposed of appropriately. Sensitive information must be destroyed. All other nonsensitive data and materials are recycled, when possible. If it is determined that a collection should be preserved through conversion to a different format and that the deteriorating format can be disposed after data and information have been migrated, the original format is discarded.

Summary

The Collections Management Plan for the U.S. Geological Survey (USGS) Woods Hole Coastal and Marine Science Center (WHCMSC) Data Library explains the records evaluation and retention policies employed by the WHCMSC Data Library, including the criteria used to evaluate records and data for inclusion into the Data Library's holdings, the scope of the Data Library's collections, and procedures for the management and operation of the library. This report also explains the roles and responsibilities of library and scientific staff.

The data librarian makes decisions pertaining to accession, retention, storage, and deaccession of all Data Library materials based on established guidelines and conditions. Raw data and associated metadata make up the majority of the Data Library's holdings. The Data Library uses field activity numbers to associate related data and support information in its sorting and storage organizational system. Research staff are responsible for providing the Data Library with raw field activity, modeling, and laboratory data and any supporting data useful in describing methodologies used during research and development activities.

The library provides a repository that serves as a resource for the local and national USGS staff, as well as outside researchers and institutions. The data collections preserved and maintained by the WHCMSC Data Library represent over 50 years of research, which would be impossible to duplicate because of time and cost constraints. These data and records continue to serve as valuable resources and provide a framework for further scientific investigations.

Selected References

Commission on Physical Sciences, Mathematics, and Applications, 1995, Study on the long-term retention of selected scientific and technical records of the Federal Government—Working papers: Washington, D.C., National Academy Press, 128 p. [Also available at http://www.nap.edu/openbook.php?record_id=9478.]

National Aeronautics and Space Administration, 2007, White paper on NASA science data retention: National Aeronautics and Space Administration Space Science Data Coordinated Archive, accessed August 4, 2014, at http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html.

National Archives and Records Administration, [n.d.], Records management by Federal agencies (44 U.S.C. chapter 31): U.S. National Archives and Records Administration Web page, accessed August 4, 2014, at http://www.archives.gov/about/laws/fed-agencies.html.

U.S. Geological Survey, [n.d.], USGS publication series definitions: U.S. Geological Survey Manual, part 1100, chap. 3, appendix A, accessed August 4, 2014, at http://www.usgs.gov/usgs-manual/1100/1100-3appendixa.pdf.

U.S. Geological Survey, 2009, Geology Discipline research records schedule: U.S. Geological Survey Manual 432–1–S5, accessed August 4, 2013, at http://www.usgs.gov/usgs-manual/schedule/432-1-s5/gd.html.
U.S. Geological Survey, 2014, U.S. Geological Survey science data exit survey form: U.S. Geological Survey Web page, accessed January 14, 2015, at http://www.usgs.gov/datamanagement/preserve/exitsurvey.php.
U.S. Geological Survey, 2015a, Preservation requirements for digital scientific data: U.S. Geological Survey Instructional Memorandum OSQI 2015–04, accessed January 14, 2015, at http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-04.html.
U.S. Geological Survey, 2015b, Review and approval of scientific data for release: U.S. Geological Survey Instructional Memorandum OSQI 2015–03, accessed January 14, 2015, at http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-03.html.
U.S. Geological Survey, 2015c, Scientific data management foundation: U.S. Geological Survey Instructional Memorandum OSQI 2015–01, accessed January 14, 2015, at http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-01.html.
University of California, 2004, Records retention and disposition—Principles, processes, and guidelines: Oakland, Calif., University of California Business and Finance Bulletin RMP–2, 30 p., accessed August 4, 2013, at http://policy.ucop.edu/doc/7020454/BFB-RMP-2.

Glossary

accession The formal archival practice of evaluating and accepting materials, data, and records for permanent retention and preservation.
catalog The process of managing and tracking materials that are uninventoried within the Woods Hole Coastal and Marine Science Center (WHCMSC) Data Library.
check-in The process of evaluating materials, data, and records before inventorying them in the Data Library collections catalog and accessioning them.
data Any assembly of observations, statistics, or analytical results used for reference or for scientific interpretation.
data library A facility that provides access to and retrieval of data and data-derived products.
deaccession The formal archival practice of removing records and collections of data from the custody and responsibility of a data library.
disposition Actions taken with regard to data and records that are no longer needed for current research or investigations, or are not appropriate for retention by the WHCMSC Data Library. Disposition encompasses the destruction of records as well as the transfer of data and records to the National Archives and Records Administration.
essential support data Any data that are neither raw nor final processed but are determined to be invaluable as a resource for understanding the nature of the raw data and aiding in their interpretation and processing.
exit survey form A form designed to capture important information from retiring or departing scientists about the nature and history of the data and records they are donating to the WHCMSC Data Library.
field activity Any activity sponsored by and tracked by the U.S. Geological Survey (USGS) Coastal and Marine Geology Program for the purpose of collecting data and making observations in the field.
field activity number A unique identifying number assigned by data managers in the USGS Coastal and Marine Geology Program to track a field activity and link it to any data and derived records to which it is associated.
final processed data Any data or derived interpretations which have been processed, vetted, and published.
intake The initial process of researching and accepting data and records into the WHCMSC Data Library.
inventory The process of documenting materials, data, and records as a part of the accession process.
inventory numbers Identifying numbers, assigned to data and records which have been accepted into the WHCMSC Data Library's collections, which are used to catalog and retrieve these data and records.
metadata Data and information used to describe a record or dataset.
minor research records Records defined as datasets that are narrowly focused, are associated with short-term tasks addressing a specific and local problem, are not part of an investigation of larger scope, and have no potential for use in future investigations.
principal investigator Primary contact and individual responsible for the planning and execution of a data collection activity.
project number An identification number assigned to data records, or collections of records, which are kept in a temporary holding and sorting area outside of the WHCMSC Data Library awaiting check-in.
raw data Any data, records, or notes that are directly collected from the field and have not been interpreted or processed in any way.

records Any items, including all books, papers, maps, photographs, machine-readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business, and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government or because of the informational value of the data in them. In the context of records preserved and maintained by the WHCMSC Data Library, these materials contain data or document data, data collections, analyses, interpretations, correspondence, background information, or historical context.

scientific authorities Researchers within the WHCMSC that are designated as points of contact to address scientific and historical questions about a given dataset or record.

secondary research records Records defined as datasets that focus on immediate, short-term tasks and are not associated with scientific investigations of larger scope.

significant research records Records defined as datasets that are irreplaceable, relevant to the USGS mission, and in a condition that allows future use.

For more information concerning this report, contact:
Director, Woods Hole Coastal and Marine
Science Center
U.S. Geological Survey
384 Woods Hole Road
Quissett Campus
Woods Hole, MA 02543-1598
WHSC_science_director@usgs.gov
508-548-8700 or 508-457-2200
or visit our Web site at:
http://woodshole.er.usgs.gov/

Publishing support by:
The Pembroke Publishing Service Center.

ISSN 2331-1258 (online)
http://dx.doi.org/10.3133/ofr20151141