RESOURCE ASSESSMENT SERVICE
MARYLAND GEOLOGICAL SURVEY
Jeffrey P. Halka, Director

COASTAL AND ESTUARINE GEOLOGY
FILE REPORT NO. 10-03

Metadata Creation for Maryland Geological Survey's Sediment Core Collections

by

Lamere Hennessee and Dale Shelton

Maryland Geological Survey
2300 St. Paul Street
Baltimore, Maryland 21218
Phone: 410-554-5500
FAX: 410-554-5502
E-mail:
lhennessee@dnr.state.md.us
dshelton@dnr.state.md.us

Partial funding for this study was provided by a grant from the U.S. Geological Survey, National Geological and Geophysical Data Preservation Program (NGGDPP)
Award Number: 09HQPA0009
Award Period: July 1, 2009 – June 30, 2010

September 2010

TABLE OF CONTENTS

Acknowledgments……...……. iv
Abstract…….. 1
Introduction……1
Background…… 2
Objectives…….. 4
The MGS Sediment Core Collections……... 4
Coastal Plain Cores (Collection P1507)…….. 5
Atlantic Continental Shelf Cores (Collection P993)……6
Chesapeake Bay Cores (Collection P1648)……6
Justification for Preserving Sediment Cores …….. 7
Metadata Creation, Conversion, and Transfer…….. …… 8
Expanding MGS's Data Preservation Database...……...… 8
Populating the Metadata Tables…….. 9
Converting Metadata Tables and Submitting Files to the National Catalog……9
Verifying the Accuracy and Completeness of the Uploaded Metadata ……10
The Data Preservation Techniques Workshop and its Aftereffects…….. 10
The MGS Sediment Core Collections: Next Steps……12
Lessons Learned……13
Collections Inventory vs. Metadata Creation...……...13
Metadata Creation and Submittal…….….. 13
The Internal Data Preservation Database……14
Labeling Collection Items …….…14
The MGS Data Preservation Advisory Panel……. 14
Changes in Workflow at MGS…….…...15
Public Access to Collections……..……15
Conclusions……... 15
References……. 15

Appendices
Appendix 1: NGGDPP Metadata Fields: Sediment Cores, Coastal Plain of Maryland (Collection P1507) ……17
Appendix 2: NGGDPP Metadata Fields: Bottom Sediment Cores, Maryland Continental Shelf (Collection P993) ……21
Appendix 3: NGGDPP Metadata Fields: Sediment Cores, Chesapeake Bay and Tidal Tributaries, Maryland (Collection P1648)……24
Appendix 4: MGS-Specific Instructions for Uploading Metadata to the National Catalog…… 28
Appendix 5: MGS Data Preservation Advisory Panel - 2009-2010 Membership……. 29

LIST OF TABLES

Table 1: Permanent collections held by MGS, by NGGDPP collection category……..… 3

ACKNOWLEDGEMENTS

We, the authors, wish to thank:

The members of MGS's Data Preservation Advisory Panel, for fostering data preservation efforts at the Survey by freely and frequently providing their thoughts and insights, based on their different realms of expertise;

Those members of the larger geoscience community who shared their opinions as to the future value of the Chesapeake Bay Earth Science Study (CBESS) samples, for making the retain/discard decision much easier;

The organizers of the July 2009 Data Preservation Techniques Workshop, for facilitating contact with other curators engaged in preserving their state's geoscience holdings and for offering new avenues and approaches to explore;

Sandy Schenck of the Delaware Geological Survey, for sharing his ideas about data preservation at the state level and suggesting possible participants on the Survey's Data Preservation Advisory Panel;

John Wilson of MGS, for sharing his invaluable knowledge of Coastal Plain cores, not least of which was their physical location at the Survey.

ABSTRACT

The Maryland Geological Survey (MGS or "the Survey") shares the concerns of other agencies and organizations engaged in geological research – that geoscience collections and data are valuable in their own right, beyond the lifetime of the projects during which they are collected or acquired, and that special efforts are required to preserve them and ensure their accessibility.

In this, its second year as a recipient of a National Geological and Geophysical Data Preservation Program (NGGDPP) grant, MGS documented each item in three of its sediment core collections and supplied the metadata to the National Digital Catalog of Geological and Geophysical Data, adding a total of nearly 4,700 records. The project served as a pilot – an opportunity for MGS to become familiar with (a) NGGDPP metadata requirements, (b) the amount of time involved in locating and compiling existing sources of information to satisfy those requirements, and (c) the process of converting metadata from an internal database to a format compatible with the National Catalog and submitting the converted files for upload. Originally, the Survey had planned to document only two sediment core collections – Coastal Plain cores and Atlantic continental shelf cores – consisting of a total of about 450 records. But, once MGS, in conjunction with its newly established Data Preservation Advisory Panel, decided to retain the sediment grab samples collected as part of the Chesapeake Bay Earth Science Study (CBESS), the Survey supplied metadata for those "cores" as well.

In the course of creating metadata for the three collections, MGS discovered that (a) the relative ease of broadly describing a collection masks the amount of work involved in compiling metadata for the items comprising the collection, (b) in a poorly organized repository, there is no substitute for first-hand knowledge in locating specific items or the sources of information needed to describe them, (c) providing metadata to the National Catalog, while time-consuming, is fairly easy to do, as long as sources of information for the required metadata can be found, and (d) changes in workflow at MGS should facilitate the documentation of new acquisitions. Finally, independently of the funded activities, MGS found that a panel of outside experts is invaluable in fostering data preservation efforts.

MGS has now completed a collections inventory and acquired experience in metadata creation – the initial steps in building what it hopes will become a first-rate repository that effectively serves the larger geoscience community in Maryland and beyond.

INTRODUCTION

The Maryland Geological Survey (MGS or "the Survey") shares the concerns of other agencies and organizations engaged in geological research – that geoscience collections and data are valuable in their own right, beyond the lifetime of the projects during which they are collected or acquired, and that special efforts are required to preserve them and, also, to ensure their accessibility.

In 2009, with funding from the National Geological and Geophysical Data Preservation Program (NGGDPP), MGS completed a broad inventory of its geoscience collections and entered information about them into the National Digital Catalog of Geological and Geophysical Data (Hennessee and Shelton, 2009). Then, over the course of one year beginning in July 2009, the Survey, in compliance with the terms of a second NGGDPP grant, created metadata for three major components of its Sediment

Core Collection:

1. Collection P1507: Intact sediment cores collected from the Maryland Coastal Plain for hydrogeologic and/or stratigraphic studies,
2. Collection P993: Intact bottom sediment cores or vertical sections of half-cores collected from the Atlantic continental shelf, primarily to identify offshore sources of sand for beach replenishment, and
3. Collection P1648: Surficial sediment (grab) samples collected between 1976-1984 from the bottom of the Maryland section of Chesapeake Bay and the lower reaches of its major tributaries as part of the Chesapeake Bay Earth Science Study (CBESS), to develop a baseline characterization of certain physical and geochemical properties of those sediments.

To fulfill the terms of this second grant, MGS made a commitment to create metadata only for the first two collections, among the best organized and documented of MGS's holdings. Doing so would enable the Survey to master the metadata creation phase of data preservation. However, in response to recommendations from the Survey's newly formed Data Preservation Advisory Panel and the larger geoscience community, MGS decided to retain the CBESS sediment samples and include them as part of this year's metadata creation activities. In all, then, MGS transferred metadata for 4,662 sediment cores to the National Catalog.

The MGS curator, the project's principal investigator (PI), attended the Data Preservation Techniques Workshop held on July 14-15, 2009, in Bloomington, Indiana.

Finally, independently of this year's grant-related activities, MGS, as indicated above, established a Data Preservation Advisory Panel of outside experts. Although formation of the Panel was not technically part of this year's NGGDPP grant, some of the findings and lessons learned presented below were informed by the existence and recommendations of the Panel.

BACKGROUND

Maryland is a relatively small, densely-populated state, with a land area of 9,844 square miles, a water area of 623 square miles, and an estimated population of 5.6 million people (MGS, 2007; U.S. Census Bureau, 2006). The state straddles six geologically diverse physiographic provinces, from the Appalachian Plateau to the Atlantic Continental Shelf, and contains an extensive network of tidal streams and bays, most notably northern Chesapeake Bay. The Atlantic Ocean forms its eastern border.

The state geological survey has been in existence since 1896. The types of geoscience collections held by MGS reflect its mission, as it has changed over the past 114 years. Current research is focused on the geological underpinnings and groundwater resources of the State. However, MGS has retained several collections from the past, when the interests of its staff and the needs of Maryland's citizenry were different than they are today. For instance, although the Survey is no longer actively engaged in paleontological research, MGS has a macrofossil collection that numbers in the hundreds of specimens. As a consequence of its longevity and the diversity of its activities, MGS possesses a wide array of holdings in a variety of formats.

MGS is still in the early stages of grappling with the long-term preservation of its data and collections in a formalized, systematic way. In 2008, NGGDPP awarded MGS a one-year grant to (1) identify and

broadly described the geoscience collections and data currently in its possession and (2) enter information about the nature, size, condition, and accessibility of those collections and data deemed "permanent" into the Collections Inventory of the National Catalog (Hennessee and Shelton, 2009). MGS identified 31 collections, 26 of which are to be held permanently. This year, MGS added a fourth sediment core collection, bringing the total to 27 permanent collections. Of the 27, eight are physical collections, and 19 are derived or indirect data collections. The distribution of the Survey's permanent collections among the NGGDPP collection categories is summarized in Table 1. Collections are housed in two buildings: the Survey's main office in Baltimore, Md., and a warehouse about 40 miles away in Matapeake, Md.

Table 1: Permanent collections held by MGS, by NGGDPP collection category.

Collection category	Permanent collections (N)
Physical Collections	
1. Auger samples	
2. Fluid samples	
3. Geochemical samples	
4. Hand samples	1
5. Ice cores	
6. Paleontological samples	1
7. Rock cores	1
8. Rock cuttings	1
9. Sediment cores	4
10. Sidewall cores	
11. Thin sections and polished sections	
12. Type stratigraphic sections	
Subtotal	**8**
Derived/Indirect Data	
13. Drilling/completion reports	1
14. Drill stem and other tests	1
15. Field notes	1
16. Geochemical data	1
17. Geophysical data	
18. Lithology logs	1
19. Maps	
20. Paleomagnetic resistivity	
21. Paper reports	2
22. Petrophysical data	
23. Photographs	3
24. Potential fields	
25. Production history	
26. Routine analysis data	2
27. Scout tickets	
28. Seismic data	1

Table 1: Permanent collections held by MGS, by NGGDPP collection category.	
Collection category	**Permanent collections (N)**
29. Source rock maturity analysis	
30. Special analysis data	
31. Stratigraphic horizons	
32. Surface and airborne data	3
33. 2-D and 3-D seismic reflection	1
34. Vertical seismic profiles	
35. Well logs	2
Subtotal	**19**
Total	**27**

Also in 2009, MGS developed a long-range data preservation plan for its non-digital holdings (Hennessee, 2009) and appointed a curator from among its scientific staff.

OBJECTIVES

Designed to further MGS's progress in data preservation, primarily through the first-time creation and submittal of NGGDPP-compliant metadata, the six objectives of the 2009 NGGDPP project, as outlined in the proposal, were as follows:

1. Assemble the information needed to develop item-by-item metadata, consistent with the metadata template, from existing internal data documentation (e.g., spreadsheets, databases, catalogs, reports)
2. Expand an internal Microsoft Access database, DataPreservation.mdb, to include not just collections, but the items that comprise them, in a format consistent with version 1.0 of the *Metadata Profile for the National Digital Catalog*:
3. Through digital transfer, provide metadata to the National Catalog for the items that comprise the proposed permanent collection at MGS.
4. Develop a long-term preservation plan for the components of the Sediment Core Collection, consistent with the Survey's *Long-Range Data Preservation Plan*.
5. Submit a final report to the NGGDPP.
6. Attend the two-day Data Preservation Techniques Workshop hosted by the Indiana Geological Survey in July 2009.

THE MGS SEDIMENT CORE COLLECTIONS

MGS holds two types of sediment cores: (1) cores collected on land, primarily from the Maryland Coastal Plain and (2) bottom sediment cores collected by boat from the Atlantic Ocean and the State's bays and lakes. While conducting its collections inventory, the Survey made the following decisions as

to the types of geologic samples that constitute "Sediment Cores," in the NGGDPP sense of the term (Hennessee and Shelton, 2009).

Generally, land-based sediment cores consist of unconsolidated sediment, which may or may not be cemented, for example, by iron or calcite. Whole cores, partial sections of core, as well as subsamples extracted directly from the cores all qualify as "Sediment Cores," as do sidewall cores taken from sedimentary rocks. Sediment drill cuttings, however, do not; rather, these are categorized as "Rock Cuttings."

Sediment cores collected from the bottoms of bays, lakes, etc. include (a) intact, whole cores, (b) intact vertical sections of sediment cores (i.e., half-cores), (c) subsamples collected from particular core intervals, and (d) surficial or grab samples collected from the top several centimeters of the surface of the bottom. MGS chose to think of the latter as broad, short (e.g., 5-20 cm long) cores collected from the top of the sediment column, instead of classifying them as "hand samples." Subsamples of cores (c) and grab samples (d) may be unprocessed or processed (e.g., the sand fraction remaining from grain size analysis of a small section of sediment core).

Over the course of the past year, MGS created metadata for three major components of its sediment core collection, each of which is described separately below.

COASTAL PLAIN CORES (COLLECTION P1507)

MGS has collected an estimated 5,000 linear feet of sediment cores from 272 wells or test holes drilled on the Maryland Coastal Plain. The Coastal Plain sediments of Maryland form a thick, clastic wedge of eastward-dipping, generally unconsolidated sediments that thicken from a few feet at the Fall Zone, the boundary between the Coastal Plain and Piedmont Physiographic Provinces, to over 7,700 feet at Maryland's Atlantic coast. These cores, some dating back to the 1950s, were collected as part of hydrogeologic or stratigraphic mapping studies of the Coastal Plain. Most of the cores were collected by MGS, many in cooperation with the U.S. Geological Survey (USGS). Others were donated to the Survey by geotechnical consultants.

The cores, 2-8 inches in diameter, were taken at depths ranging from the land surface to thousands of feet deep. Some represent attempts to take continuous cores for the entire depth of a particular test hole. In other test holes, only a single two-foot core may have been retrieved. Many of the cores have been analyzed for permeability and for paleontological, mineralogical, textural, and/or various other types of data. Many of the cores are described, at least in part, in Survey publications.

Cores are stored mainly in a basement storage area of the Survey's main building in Baltimore. Because of the condemnation and closure of the Survey's former storage facility, located a few blocks from the main building, the core collection was hastily moved several years ago to its current location. No attempt was made to note the shelf numbers of the individual core boxes when they were relocated. Nor are the cores stored in a logical fashion for ease of access. Many need to be repackaged and/or relabeled.

A catalog of the Coastal Plain core collection is maintained by the Survey's Hydrology and Hydrogeology Program. The catalog, however, is out of date and, as indicated above, contains no information as to the specific physical location of a core. Also, information on the condition and actual

footage of cores should be added to the catalog. On a positive note, most cores are associated with a uniquely identified well or test hole with known latitude and longitude.

ATLANTIC CONTINENTAL SHELF CORES (COLLECTION P993)

This collection consists of 282 intact, continuous vibracores retrieved from the Atlantic continental shelf offshore of Maryland, between the Delaware State Line (latitude 38°28' N) and the Virginia State Line (latitude 38°00' N) and eastward approximately 30 km (longitude 74°45' W) to a water depth of 30 m. The cores were collected between 1984-1997 as part of several studies involving cooperative agreements between the States of Maryland and Delaware and various federal agencies. The objectives of the investigations varied: define the Quaternary stratigraphy of Maryland's inner continental shelf; assess the economic potential of offshore sand resources, based on heavy mineral analysis; identify offshore sources of sand suitable for beach replenishment in Ocean City, Md., and other areas along the Delmarva coast; assess the feasibility of various projects addressing environmental and navigational problems stemming from the stabilization of the Ocean City Inlet.

With a grant from the Minerals Management Service (MMS) in the mid-1990s, the Survey established a repository of Maryland continental shelf cores at its Matapeake facility. During post-collection processing, most cores, initially 3 5/8" in diameter, were cut in half length-wise. Half of each core was sampled for a variety of laboratory analyses, and the other half was archived. Archived sections, up to 1.5 meters (5 feet) in length, were sealed in polyethylene plastic sleeves, labeled with a researcher-assigned ID and the year of collection, and stored on metal shelves especially designed for the purpose. Since then, the cores have been stored under-roof, but at ambient outdoor temperature and humidity.

In addition to preparing the cores for long-term storage, MGS created a detailed catalog of the collection, in paper and digital formats (Wells and Conkwright, 1996). These were the primary sources of information used to create collection metadata, along with conversations with the catalog authors.

CHESAPEAKE BAY CORES (COLLECTION P1648)

About thirty years ago, between 1976 and 1984, the States of Maryland and Virginia, in cooperation with the U.S. Environmental Protection Agency (EPA), undertook a multidisciplinary research effort designed in part to address the managerial and environmental problems of toxic substances in Chesapeake Bay. One of the main objectives of the Chesapeake Bay Earth Science Study (CBESS) was to determine the spatial distribution of various physical and chemical properties of bay bottom sediments – grain size, water content, carbon content, and sulfur content – which, in turn, affect the distribution of toxic substances in the estuary. For its part, MGS collected 4,255 surficial sediment (grab) samples from the Maryland section of Chesapeake Bay and the lower reaches of its major tributaries and analyzed them for the four constituents of interest.

In accordance with the terms of the EPA grant, MGS archived the original wet sediment samples left over from lab processing, with the intent of keeping them for 30 years. Samples were packaged in labeled plastic bags and placed inside sealed and labeled glass jars. The glass jars were then stored in labeled cardboard boxes on shelves at the Survey's warehouse/garage in Matapeake. The building is insulated and wired for electric lights, but temperature and humidity cannot be regulated, except by opening or closing doors.

MGS did not include the CBESS sediment cores in this year's NGGDPP proposal. At the time the proposal was written, Survey staff had yet to decide whether the collection should be permanently retained. The arguments for keeping the samples were that (a) they represent a snapshot of the entire Maryland section of the Bay at a particular point in time and (b) they were acquired at considerable effort and expense. The main argument against retention was that, for 30 years, the samples have been stored at ambient outdoor temperatures. At a minimum, their geochemical properties have changed. (At this point, storage space is not an issue for MGS.)

The matter was discussed at the first meeting of the Survey's Data Preservation Advisory Panel. Panel members were unsure about the future value of the CBESS samples to geoscience researchers or educators and suggested that MGS poll other members of the regional geoscience community. In addition to the members of the Advisory Panel and the staff of the Survey's Coastal & Environmental Geosciences Program, MGS contacted the heads of several geoscience programs at Maryland and Virginia colleges and universities, the directors of the federal (EPA and NOAA) Chesapeake Bay Program, state government employees at the Maryland Department of Natural Resources, members of the Baltimore Ecosystem Study, and a private foundation devoted to public education and restoration of the Chesapeake Bay. Of the 22 outside people contacted (non-MGS, non-Advisory Panel), five responded, though some polled their colleagues before replying.

The Survey received a range of responses. The strongest arguments against keeping the samples were their age and the fact that they have been stored at ambient outdoor temperatures. Geochemically, "the age of the samples is problematic," in the opinion of one sedimentary geochemist. One respondent wavered, "My heart wants to say keep them, but my head says they can go away." However, faculty members at two branches of the University of Maryland both felt strongly that the "samples represented a valuable resource, with the potential to be used for research in the future," for example, "to look at some kind of tracer we don't even know about sometime in the future." Both urged MGS to "keep this valuable collection for the foreseeable future." Furthermore, if space were an issue, one branch of the university offered to look into storing the samples there.

After considering the matter in-house, MGS decided to keep the samples, at least for now, based on the fact that space is not a pressing issue and that two respondents thought the collection might one day be valuable.

Like the other two sediment core collections, the CBESS samples are well-documented, although the information about them is scattered: a final project report, a digital database of lab results and associated FGDC-compliant metadata, a dilapidated lab notebook indicating the numbers of the jars in which samples are stored, and field notebooks and associated maps. Fortunately, one of the authors was actively involved in that project and knew where to find these sources.

JUSTIFICATION FOR PRESERVING SEDIMENT CORES

Preservation of these sediment cores is important for a number for reasons. Cores may be useful for purposes other than those for which they were originally collected. For instance, as new analytical techniques become available, they could be applied to existing cores, without the added expense of collection. In particular, the sediment cores from the continental shelf are now of interest for assessing marine habitat and, with the possible advent of offshore wind farms, for siting windmills.

Cores are expensive to collect. MMS has expended an estimated $500,000 for the collection and analysis of Atlantic shelf cores, a dollar figure matched by MGS. In 2008, the cost of drilling 150-ft deep test holes in Maryland Coastal Plain sediments, based on a proposal submitted by the winning bidder, was $20/ft for split-spoon coring, $24/ft for drilled footage, plus a rig-mobilization charge of $1,200 per site (Wilson, J., pers. comm.).

Cores may be unique and/or impossible to replace. CBESS samples, collected according to a dense, one-kilometer grid, represent the bottom sediments blanketing the entire Maryland portion of Chesapeake Bay at the end of the 20th century. Some of the Atlantic shelf cores represent undisturbed ocean bottom sediments, prior to the mining of sand for beach nourishment. As for Coastal Plain cores, Maryland is the country's fifth-most densely populated state, and the state's planning office is predicting a 27% increase in population by 2030 (Rein, 2009). With development proceeding so rapidly, it may be impossible to retrieve Coastal Plain cores from sites that MGS has already sampled. Coastal Plain geologists are already finding it harder and harder to locate new drilling sites (Wilson, J., pers. comm., 6/15/2010). Retaining those cores may allow future geologists to make judgments or interpretations with more information than they might otherwise have.

Cores may facilitate the assessment of natural resources or geologic hazards. The latter particularly may require quick answers, over a timeframe that does not allow for additional data collection.

METADATA CREATION, CONVERSION, AND TRANSFER

This year's project can best be described as a pilot that enabled MGS to master the metadata creation phase of data preservation, including digital metadata transfer to the National Catalog. Meeting that goal entailed (a) expanding an in-house data preservation database to include, for the three chosen collections, item-specific metadata, (b) populating the new metadata tables, (c) converting the tables to a format acceptable to the National Catalog, (d) submitting the reformatted files, and (e) verifying that the uploaded metadata records correspond with MGS's internal records in terms of accuracy and completeness.

EXPANDING MGS'S DATA PRESERVATION DATABASE

Last year, MGS created a Microsoft Access database, DataPreservation.mdb, as an internal digital catalog of its collections. The structure of that database was documented in an appendix to last year's report to the NGGDPP (Hennessee and Shelton, 2009). This year, the database was expanded to include three metadata tables, one for each of the collections selected for documentation. The three tables include all of the optional and mandatory metadata fields suggested or required by the NGGDPP, as well as additional fields, some of which were concatenated to populate certain NGGDPP fields.

One of the problems that MGS encountered in this step involved the structure of the database, the design of which has been driven in large part by the requirements of the NGGDPP. Theoretically, because the same metadata are required of all collection items, one large metadata table should suffice. However, the information used to describe sediment cores is very different from that used to describe other collections, such as maps. It seemed that the best way to deal with those differences was to build separate tables for each collection and then extract the information to populate an NGGDPP metadata table from those.

Then the question became, "How many sediment core collections should MGS recognize?" Dividing the sediment cores into two collections, one consisting of land-based cores and the other of water-based cores, seemed reasonable. But should the sediment cores collected from the Atlantic Ocean be grouped with the grab samples collected as part of an entirely different project, with different objectives, from the Chesapeake Bay? In the end, it was easier to create the metadata for the two water-based core collections separately, so that is what MGS did. In conversations with NGGDPP's Frances Pierce, the rationale became one of recognizing unique study areas. That is, researchers interested in open ocean sediments might find it useful for such cores to be maintained in a different collection than those retrieved from an estuarine system.

POPULATING THE METADATA TABLES

Although it required populating more than one metadata table to recognize a problem and arrive at a solution, MGS eventually decided to create a form for each collection with the appropriate metadata information, as well as information sources, explanations, and examples, as appropriate. The completed form will serve to guide the Survey curator in the future, when he or she adds new items to a previously documented collection. The idea is similar to one suggested in the NGGDPP instructions, *Preparing Metadata for the National Digital Catalog* (05/15/2009), which provides a worksheet for mapping existing digital data into the metadata fields. The completed forms for the three sediment core collections can be found in Appendices 1-3.

Because so much of the marine and estuarine research done at the Survey is project-based, MGS decided to include the project acronym as part of the metadata field "Title" and the full name of the project in the "Abstract" field for the two water-based sediment core collections. That will simplify searching by project. It will also result in unique titles, in a discipline where, so many times, the first sample collected as part of a project is assigned an ID of "1."

CONVERTING METADATA TABLES AND SUBMITTING FILES TO THE NATIONAL CATALOG

Approaching this step with some trepidation, the authors were pleased and surprised to discover that their fears were unfounded. Converting the Access metadata tables to .csv-formatted files and uploading those files to the National Catalog were easy, due mainly to the clear instructions in *Preparing Metadata for the National Digital Catalog* (05/15/2009). (Familiarity with the requirements of developing FGDC-compliant metadata was also helpful.)

The two-stage Validate-Request Load process worked very well. MGS particularly appreciated the human link in the chain. Of the two problems that MGS encountered in uploading files, the NGGDPP's Rick Brown caught one between the two stages of the process. In one of its files, MGS had supplied a single alternateTitle comprised of four bits of information separated by commas, not realizing that commas, in this case, would result in four separate, stand-alone titles. Rick spotted the potential problem immediately and recommended substituting semi-colons for the commas. The second problem, also easily resolved, was due to MGS's having misspelled the name of one of the NGGDPP fields in the upload file (datasetReferenceDat**e** is NOT spelled datasetReferenceDat**a**).

Detailed, MGS-specific instructions for accomplishing this step are included in Appendix 4.

VERIFYING THE ACCURACY AND COMPLETENESS OF THE UPLOADED METADATA
Finally, MGS verified the completeness and accuracy of the metadata upload. In terms of completeness, MGS checked that the total number of records in its internal metadata tables (one per collection) matched the number uploaded to the National Catalog. Then, for 20 records in each collection, MGS verified the accuracy of the uploaded information, that is, MGS verified that the information in the National Catalog matched the information in (a) the internal metadata tables and (b) the source documents or files from which the metadata were compiled. The Survey chose the checked records by dividing the total number of records in a collection by 20 and, beginning with 1, adding the quotient to each successive selection. For example, for the 125 Coastal Plain cores, MGS verified the accuracy of the first and every sixth record after that (i.e., records 1, 7, 13, 19, etc.). No errors were detected.

After verifying the accuracy and completeness of the metadata upload, MGS reviewed the associated information contained in the Collections Inventory. In some cases, the number of items uploaded differed from the initial estimates reported in the original description of a collection. Also, as a collection grows, the contents of the collection may be either broader or more restrictive than the original description of the collection indicates.

THE DATA PRESERVATION TECHNIQUES WORKSHOP AND ITS AFTEREFFECTS

The MGS curator attended the two-day Data Preservation Techniques Workshop sponsored by the Illinois Geological Survey in July 2009 and documented her immediate reactions to the proceedings in last year's report to the NGGDPP:

> "At the Data Preservation Workshop, representatives of several mid-Atlantic state surveys agreed that, given the funding and staffing problems each faced, it seemed inefficient for each survey to designate a curator, develop preservation strategies, and maintain separate repositories. There might be economies of scale in cooperating in a regional repository. Such a repository might also serve the needs of other state agencies and private companies that collect geologic materials…
>
> …it was apparent that, in terms of data preservation, state geological surveys can be divided into the experienced and the inexperienced. The former are already well along the path to establishing respected geoscience repositories. The latter are just beginning the process. As one of the inexperienced, MGS has benefited enormously from the work of the NGGDPP and its predecessors…" (Hennessee and Shelton, 2009)

For MGS, the most consequential effect of the workshop was persuading the curator of the need to form a Data Preservation Advisory Panel of outside experts and to define "expert" broadly. This year, independently of the NGGDPP project, but consistent with that Program's guidelines, MGS formed such a panel. At the recommendation of another state survey attendee at the Data Preservation Workshop, the MGS curator apprised the Maryland Geologic Mapping Advisory Committee (GeoMAC) of data preservation activities at the Survey and invited members of that Committee to serve on the Data Preservation Advisory Panel. Many GeoMAC members agreed to do so. In addition to GeoMAC, MGS

invited a number of other geoscientists named by MGS staff during the collections inventory as being familiar with particular collections. Representatives of several other disciplines - a librarian, an archeologist, and two archivists – also agreed to participate on the Panel (see Appendix 5 for the Panel membership list).

The Panel met for the first time in November 2009 and formulated a list of the group's responsibilities:

- Advise MGS on the disposition of its collections (assist in establishing criteria for keeping or disposing of items),
- Help MGS prioritize the order in which collections will be documented, organized, and preserved,
- Double as a "User Committee" by suggesting how MGS might improve access to archived collections,
- Suggest possible sources of outside funding or partnerships in support of the repository, and
- For those Panel members familiar with library, museum, or archives "best practices," share their expertise with MGS.

With the structure of the GeoMAC in mind, MGS requested that a non-Survey member serve as the chair of the Advisory Panel. That way, if the Panel endorsed a collections-related proposal submitted by MGS, there would be no obvious conflict of interest. The Deputy State Archivist of the Maryland State Archives (MSA) offered to serve as chair and was immediately elected.

At the same meeting, Panel members discussed the disposition of the CBESS samples. They recommended that, before making a final decision, MGS contact local universities and other agencies suggested by Panel members to discover (a) if the samples were still valuable, given their age and long-term storage at ambient outdoor temperatures, and (b) if the agency polled had the space and inclination to store the samples if MGS decided to dispose of them. In response to that polling, MGS decided to retain the CBESS samples as part of its permanent sediment core collection.

The Advisory Panel has also been supportive of two proposals that MGS submitted to further its data preservation activities: the FY2010 NGGDPP grant to create metadata for five more MGS collections and a National Historical Publications and Records Commission (NHPRC) grant, submitted in conjunction with the Maryland State Archives, to scan the entire collection of MGS-produced maps – an estimated 1,000 maps and oversized publication inserts – and post it online. In both cases, the Panel wrote a letter to the grant agency endorsing the proposal.

MSA representatives on the Advisory Panel arranged a meeting between the Archives and the Survey to help MGS formulate a plan for relocating two of its paper-based collections (maps and reports) from an Archives Room lacking proper climate control to another room in the Survey's main building with a more reliable air conditioning unit. Both MSA and the librarian on the Advisory Panel lent MGS carts and dollies, which greatly facilitated the move. Not only did the librarian design a floor plan for the new Archives Room, but she also assisted in the move itself.

Incidentally, following a casual conversation with the librarian during the course of the move, MGS has begun working cooperatively with her to scan the Survey's many hardcopy publications. And in response to another of her suggestions, MGS has established a working partnership with the Johns Hopkins University (JHU) Sheridan Libraries to preserve and scan a subset of the Survey's maps, which will be made available to the public through JHU's JScholarship website.

THE MGS SEDIMENT CORE COLLECTIONS: NEXT STEPS

In an effort to meet the requirements of this year's grant, MGS chose to use only existing "written" documentation about sediment cores in creating and submitting metadata records to the National Catalog. Because of time constraints, the Survey did not verify the location of each sediment core for which a metadata record was provided. Furthermore, MGS had planned to reorganize, repackage, and relabel the items comprising its sediment core collections as part of this year's activities. The Survey encountered two obstacles in completing those activities: (a) the lack of funding for packaging materials, and (b) the decision to retain the CBESS samples and provide related metadata to the National Catalog. Although the latter enabled MGS to add almost 4,700 metadata records to the National Catalog, instead of the proposed 450 or so, doing so was quite labor intensive.

The following are the logical next steps in the preservation of the Survey's three sediment core collections:

- Develop a plan for reorganizing the cores, preferably relocating them to a single storage area. Currently, Coastal Plain sediment cores are scattered, stored in several different rooms in the main building in Baltimore and stacked in such a way that accessing one core box requires moving many others. Ideally, all sediment cores would reside in a single location and be organized and stored in such as way as to make them easy to find and access.
- If cores remain at the Matapeake facility, improve storage conditions, including climate control; ensure that field equipment kept at the site does not interfere with core storage or access.
- Match each core in a collection to its corresponding metadata record and add the exact location of the core (building room number, shelving unit, shelf) to the record; upload changes to the National Catalog.
- Develop internal protocols for the information that should consistently be included on core box labels, as well as a plan for labeling cores that could be expanded to other collections.
- Repackage and relabel core boxes as needed, making sure that all required information is included and that the information is correct (e.g., no two boxes contain the same interval of the same core).
- For newly collected cores that are to be held permanently, implement a change in workflow so that required NGGDPP metadata are collected routinely by Survey researchers and forwarded to the curator as soon as possible after a core is collected.
- As a related activity, determine how soon after collection a core must be listed in the National Catalog (i.e., the number of years after collection).
- Modify the internal database design so that (a) sediment core tables contain content-based information for web-based delivery, in addition to NGGDPP metadata fields and (b) items collected as part of same project but belonging to different collections are linked.
- Develop protocols for the use of the collections by outsiders. Begin thinking about how to make the collections more accessible to outside users and how to encourage their use (e.g., advertise the existence of collections; provide instructions for their use).
- If possible, locate additional metadata for the cores lacking all of the required metadata, which are, therefore, not yet listed in the National Catalog.
- Explore entering data related to water-based sediment cores into a national database, such as NOAA's Index to Marine & Lacustrine Geological Samples.
- Discuss ideas to fund the repository with the Survey's Data Preservation Advisory Panel

LESSONS LEARNED

COLLECTIONS INVENTORY VS. METADATA CREATION

A general inventory of collections masks the amount of work needed to itemize an existing collection. Last year, estimating the number of items comprising a collection seemed relatively straightforward. However, locating and describing each item turns out to be something else entirely. At MGS, the items comprising a collection may be scattered throughout the building. So, really, a shelf-by-shelf inventory of all of MGS's holdings is needed. Only then will the actual location of all items in any one Survey collection be known. Once that is done, then the process of organizing and centrally storing a collection can begin.

Having said that, the process of top-down documentation of collections, with increasing granularity in the descriptions, makes good sense in terms of coming to grips with the entirety of the Survey's holdings. Clearly, the elephant can only be eaten one bite at a time, and the collection-inventory-followed-by-metadata-creation approach is a satisfactory way to accomplish the task. It is just more complicated than it sounds.

METADATA CREATION AND SUBMITTAL

The fact that "catalogs," rather broadly defined, already existed for the three sediment core collections made it relatively easy to provide metadata for them, particularly when the catalogs were in a digital format. It was also useful that staff members were quite knowledgeable about the collections and could provide or serve as sources of information. Without the catalogs or the personal expertise, creating metadata would have been extremely difficult, if not impossible.

Because MGS has heretofore not standardized a process for collecting and storing metadata, obtaining the information used to create metadata for existing collections requires consulting the reports and, ideally, the researchers associated with particular projects. Because publications and personnel vary from project to project, each collection requires documentation as to the source(s) of information used in creating the associated metadata, as well as the specific content to be included in each field. This is the idea of the metadata template outlined in *Preparing Metadata for the National Digital Catalog* (05/15/2009) for mapping information from an internal database to the expected NGGDPP format. To implement the idea at MGS, where not all sources are digital, the Survey developed a standard form for describing the sources and contents of a collection's NGGDPP metadata fields (see Appendices 1-3). These forms may have to be modified if the items comprising a collection are collected during the course of multiple projects.

It is unclear from NGGDPP publications whether geographic coordinates are to be reported as or adjusted to the North American Datum of 1983 (NAD 83). If that is not a standard requirement, then the agency reporting those coordinates should specify the horizontal datum, for example, in the metadata field "alternateGeometry."

In terms of verifying the accuracy of metadata upload, it probably makes more sense to check a certain <u>percentage</u> of records in a collection, rather than a certain <u>number</u>. Checking 20 records in a collection of 125 results in a verification of 16% of the records, compared to <1% in a collection of 4,255 records.

THE INTERNAL DATA PRESERVATION DATABASE

As reported last year, MGS has still not completely resolved the issue of separating the highly integrated products collected during the course of a particular project into separate collections. For example, it is not unusual for the Coastal & Environmental Geosciences Program to collect a sub-bottom seismic profile and, from that seismic information, to determine core sampling locations. Once a core is collected, it is typically x-rayed, split longitudinally, photographed, described lithologically, and subsampled for textural and geochemical analyses. So, the physical core is associated with a whole suite of derived products: seismic runs, x-rays, photos, and analytical results. The PI is accustomed to keeping all of the derived products associated with a project, in this case, a set of sediment cores, together. Separating them into collections of seismic data, photographs (both x-rays and photos), lithology logs, and routine analysis data, introduces a sense of disorder – dispersing a group of products that logically belong together – and, consequently, requires that some provision be made for establishing a connection among all of a project's physical and derived products. MGS plans to meet that need via its internal database but has not yet modified the database design to do so.

LABELING COLLECTION ITEMS

To the extent possible, MGS needs to develop a standard means of labeling items in a collection, for ease of locating those items. The CBESS samples provided a worst-case example of the failure to do so. In designing that project's sampling scheme, MGS researchers divided the Maryland part of the Chesapeake Bay into several dozen "tiles," based on a series of Oyster Bar Charts, numbered 0-36 (with gaps). Throughout processing and storage, samples remained grouped by those chart numbers. At the time of collection, a now-obsolete coordinate system, consisting of Raydist Red and Green lanes, was used to locate samples in the field. Upon collection, samples were labeled with the Oyster Bar Chart number and the date of collection and were then consecutively numbered, beginning anew each day with "1." To further complicate matters, two sampling schemes were developed, one for shallower "nearshore" samples and the other for deeper water "mid-bay" samples. To retrieve a particular sample, the curator must know the following, in order: (1) sample collection technique (mid-bay/nearshore), (2) Oyster Bar Chart number, (3) Raydist Red and Green lanes, (4) jar number, and (5) collection date and sample number. Unfortunately, the Oyster Bar Chart number and the jar number were not included in the spreadsheet of CBESS data. However, MGS had retained the lab notebook used to track sample processing and post- processing storage in glass jars, the labels of which were noted in the notebook. Adding that information (Oyster Bar Chart and jar numbers) to the metadata table entailed matching each entry in the spreadsheet with its corresponding entry in the lab notebook, a long and tedious process. Standardized item labeling would eliminate such problems.

THE MGS DATA PRESERVATION ADVISORY PANEL

Although the curator was initially skeptical of the value of a committee of outside experts, the MGS Data Preservation Advisory Panel has been one of the most significant developments in the Survey's data preservation efforts to date. The Data Preservation Techniques Workshop at Indiana University (IU) convinced her of the need for such a panel. A visit to the IU archival library facility and the workshop presentation by a librarian showed the value of input from people whose (non-geological) profession it is to care for and manage collections. Also at the workshop, the collections manager from neighboring Delaware suggested a possible approach for garnering membership. From its inception, the Advisory Panel has fostered data preservation at MGS: helping to resolve thorny questions (e.g., keep or discard a particular collection), endorsing proposals, forming partnerships in applying for preservation-related grants. Furthermore, the thoughtful input of outside experts, who take their responsibilities

seriously, somehow legitimizes the entire archival effort. The very existence of the Advisory Panel makes it a little more difficult to halt the process (e.g., reassign the curator to other duties), and it holds MGS accountable to a group of people who are giving freely of their time for the sole purpose of helping the Survey to do a better job. And then there is the serendipity effect – if people talk, one thing often leads unpredictably to another.

CHANGES IN WORKFLOW AT MGS

MGS clearly needs to formulate and institute a system for (a) determining whether or not a collected item should be permanently retained, (b) reporting minimum metadata (i.e., NGGDPP required metadata) for such items to the curator as soon after collection as possible, and (c) selecting a consistent maximum lag time between collecting an item and uploading the associated metadata to the National Catalog. This will require a change in workflow at the Survey. However, instituting these changes will facilitate future preservation activities.

PUBLIC ACCESS TO COLLECTIONS

MGS must develop protocols, including agreements, for requesting and utilizing items in its collections. Over the course of the past year, researchers from the USGS requested a small sample of seldom-penetrated basement rock from one of the MGS sediment cores. Only one MGS staff member knew the whereabouts of the core, and he provided the sample to the USGS researchers without involving the Survey curator. He asked nothing of the researchers in return, such as providing MGS with a final report of their findings or requesting that they acknowledge MGS as the source of the sample.

CONCLUSIONS

During the past year, MGS has successfully completed a number of activities in building what it hopes will become a first-rate repository that effectively serves the larger geoscience community in Maryland and beyond. Having created and uploaded metadata for its three sediment core collections, MGS now has a better understanding of (1) the information required by NGGDPP to document the items that comprise a collection, (2) the potential difficulties in locating and compiling the sources of such information for legacy collections, (3) the process for converting metadata from an internal database to a format compatible with the National Catalog, and (4) the need to adjust workflow at the Survey so that researchers provide all required metadata to the curator as soon as possible after an item is collected. Inspired by the Data Preservation Techniques Workshop, MGS has established an external Data Preservation Advisory Panel. This year, Panel members advised the Survey in an important keep-or-discard-a-collection decision, endorsed two data preservation proposals, formed a partnership with MGS in submitting one of those proposals, and assisted in scanning MGS publications and relocating the Survey's Archives room. In its data preservation efforts, the Survey's next steps are to continue documenting its remaining collections, to seek funding for and prepare all of the collections for long-term preservation, and to begin addressing public access to the collections.

REFERENCES

Hennessee, L., 2009, Long-range data preservation plan: Year 1: Baltimore, Md., Maryland Geological Survey, Coastal and Estuarine Geology File Report No. 09-04, 42 p.

Hennessee, L., and Shelton, D., 2009, Inventory of geological data and collections at the Maryland Geological Survey: Baltimore, Md., Maryland Geological Survey, Coastal and Estuarine Geology File Report, 43 p.

Kerhin, R.T., Halka, J.P., Wells, D.V., Hennessee, E.L., Blakeslee, P.J., Zoltan, N., and Cuthbertson, R.H., 1988, The surficial sediments of Chesapeake Bay, Maryland: Physical characteristics and sediment budget: Baltimore, Md., Maryland Geological Survey Report of Investigations No. 48, 82 p.

Maryland Geological Survey (MGS), 2007, Fact Sheet 2: Land areas, inland-water areas, and length of shorelines of Maryland's counties, http://www.mgs.md.gov/esic/fs/fs2.html, [3/31/2008].

Rein, L., 2009, O'Malley outlines growth strategy: bills proposed to preserve farmland, limit Bay pollution, *Washington Post*, Tues., Jan. 13, 2009, p. B04.

U.S. Census Bureau, 2006, 2006 population estimates, United States – states; and Puerto Rico, http://factfinder.census.gov/, [3/31/2008].

Wells, D., and Conkwright, R., 1996, Physical inventory and repository of vibracores collected on Maryland's Continental Shelf: Baltimore, Md., Maryland Geological Survey, Coastal and Estuarine Geology Program File Report No. 96-6, 65 p.

APPENDIX 1
NGGDPP Metadata Fields:
Sediment Cores, Coastal Plain of Maryland
(Collection P1507)

Sources of Information:

- List of wells/cores, by county, prepared by MGS's John Wilson for the 2009 NGGDPP proposal
- U.S. Geological Survey's National Water Information System (NWIS) website

MetadataID

Definition: Metadata identification number
Value: 1…
Source: Assigned automatically by Microsoft Access

CollectionID

Definition: NGGDPP collection identification number
Value: P1507 (Sediment Cores, Maryland Coastal Plain)
Source: DataPreservation.mdb – tblCollection – field "USGSCollectionID"

Title

Definition: Official, human-readable title for individual record, used in listings & search results (short, distinctive) – mandatory
Value: the well number, as reported in the online USGS National Water Information System (NWIS) database; for counties, usually in the form:

XX Yy # OR
XX Yy ## OR
XX Yy ###

Where XX is the two-character county abbreviation:

County	Abbreviation	County	Abbreviation
Allegany	AL	Kent	KE
Anne Arundel	AA	Montgomery	MO
Baltimore	BA	Prince George's	PG
Calvert	CA	Queen Anne's	QA
Caroline	CO	St. Mary's	SM
Carroll	CL	Somerset	SO
Cecil	CE	Talbot	TA
Charles	CH	Washington	WA
Dorchester	DO	Wicomico	WI
Frederick	FR	Worcester	WO
Garrett	GA	Montgomery	MO
Harford	HA	Prince George's	PG
Howard	HO		

Yy is ?
or ## or ### is the one- to three-digit number assigned to the well

NOTE: In the Title, spacing and capitalization matter.

Source: MGS well catalog
Example: AA Bf 100

Alternate Title

Definition: Additional title identifiers for individual record (e.g., for further identification by other Web service interfaces); textual titles or specific sample IDs used by collection – optional
Value: "Site number" from the USGS National Water Information System (NWIS) database
Explanation: Concatenation of latitude and longitude of well location, in degrees, minutes, and seconds, with another two digits appended to the end
Source: USGS National Water Information System (NWIS) database
Example: 384332076322401

Abstract

Definition: Human-readable description of individual record, used to help determine nature of underlying physical data resource; contains much information about data resource – mandatory
Value: Includes, at a minimum:

- County in which core is located, with more specific site location information if available
- Number and depth boundaries of core sections
- Indication of presence of basement rock at bottom of core
- Depth of hole
- Altitude of land surface at well location, in feet, along with vertical datum
- Expand to include physical location of core at MGS (e.g., "Main building, Room xxx, Unit xxx, Shelf xxx, Box xxx")

Examples:

(a)
Sediment core collected from Waugh Chapel, Anne Arundel Co., Md.; includes basement core at bottom
Hole depth = 1025 ft
Altitude of land surface = 150 ft (NGVD29)

(b)
Sediment core collected from Anne Arundel Co., Md.; two sections of core between 20-347 ft & 282-468 ft
Hole depth = 468 ft
Altitude of land surface = 50 ft (NGVD29)

SupplementalInformation

Definition: Information on how to access physical data represented by metadata record (e.g., general for entire collection, such as URL, or specific reference to online resource, like ordering system with specific ID) - mandatory
Value: "Contact the Program Chief of the Hydrogeology & Hydrology Program at the Maryland

Geological Survey, (410) 554-5500"
Source: n/a

Coordinates

Definition: Geographic coordinates (longitude, latitude), in decimal degrees – mandatory
Value: (-)decimal longitude, decimal latitude
Source: Latitude and Longitude from the USGS National Water Information System (NWIS), converted from degrees, minutes, and seconds to decimal degrees

AlternateGeometry

Definition: Alternate method of storing geospatial footprint; description of authoritative source of geographic location & how simple coordinates derived – optional
Value: "Coordinates are based on the North American Datum of 1927"
Source: USGS National Water Information System (NWIS)

OnlineResource

Definition: URL pointer(s) to textual information about specific record - optional
Value: http://waterdata.usgs.gov/nwis/
Source: n/a

BrowseGraphic

Definition: URL pointer(s) to images representing specific record - optional
Value: none supplied
Source: n/a

Date

Definition: Meaningful date (e.g., collection date) attached to record; may be to any degree of precision or left blank (e.g., 20010101, 1939-1945, -20030331, 2000-) - optional
Value: date of core collection; generally not supplied
Source: MGS well catalog

DatasetReferenceDate

Definition: Reference date indicating currency of underlying data record (e.g., date metadata record added to National Catalog); format=YYYYMMDD - mandatory
Value: Date record provided to NGGDPP for uploading to National Catalog
Source: Provided by curator

VerticalExtent

Definition: Vertical extent (e.g., vertical depth information for rock core samples); contains 2-3 elements: unit of measure, max value, min value (e.g., m, 35.4, 0 => rock core measured at 35.4 total meters)
Value: Maximum value is either 0 (land surface) or the upper boundary (depth) of the topmost section of core; minimum value is either the lower boundary (depth) of the bottommost section of core or the depth of the hole
Source: MGS well catalog; USGS National Water Information System (NWIS) database
Example: feet, 103, 503

Location of Archived Cores

Cores are stored in the Survey's main building in Baltimore, in the main building basement, the annex basement, the annex hallway, and probably elsewhere

Additional Information about the Sources, Samples, Etc.

To access the NWISWeb Water Data website:

1. Go to the USGS website www.usgs.gov
2. From the "Science Areas" panel, choose "Water" =>
3. Under "WATER DATA FOR THE NATION - National Water Information System (NWIS)," choose "Maryland" from the drop-down menu, "Data by State…"
4. From the "USGS Water Data for Maryland" screen, choose the Data Category "Site information"
5. On the "USGS Water – Data Site Information for Maryland" screen, click "Site information"
6. From the screen "Site Inventory for Maryland – Choose Site Selection Criteria," make the following selections:

   Site Location = County
   Site Identifier = Site Name
   Site Attribute = Site Type
   <Submit>

   From the next screen:

   Site Name = Aa A (must supply at least 3 characters)
   County = Anne Arundel, for example
   Site type = Well

   Choose Output Format (be sure to click on circular button in front of these options)

   Table of sites sorted by Site name grouped by County
   Site-description information displayed in tab-separated format, saved to file

APPENDIX 2
NGGDPP Metadata Fields:
Bottom Sediment Cores, Maryland Continental Shelf
(Collection P993)

Sources of Information:

- Published catalog of vibracores collected on Maryland's Continental Shelf (Wells and Conkwright, 1996)
- MGS Microsoft Excel spreadsheet, MMS_COREINFO.xls

MetadataID

Definition: Metadata identification number
Value: 1…
Source: Assigned automatically by Access

CollectionID

Definition: NGGDPP collection identification number
Value: P993 (Bottom Sediment Cores, Maryland Continental Shelf)
Source: DataPreservation.mdb – tblCollection – field "USGSCollectionID"

Title

Definition: Official, human-readable title for individual record, used in listings & search results (short, distinctive) – mandatory
Value: Project acronym <comma> <space> "Core" <space> Researcher-assigned Core ID
Source: Excel spreadsheet, MMS_COREINFO.xls, fields "Project" and "CoreID"
Example: "OCBRP_I, Core 1-1"

Alternate Title

Definition: Additional title identifiers for individual record (e.g., for further identification by other Web service interfaces); textual titles or specific sample IDs used by collection – optional
Value: none supplied
Source: n/a

Abstract

Definition: Human-readable description of individual record, used to help determine nature of underlying physical data resource; contains much information about data resource – mandatory
Value: Description of core, including:

- Type of core (e.g., vibracores, gravity core, grab sample)
- Agency that collected the core
- Full name of project under which core was collected
- General purpose of project (i.e., the reason why the core was collected)
- Other information/items associated with core(e.g., specific lab analyses, seismic records, core x-rays)
- Expand to include physical location of core at MGS (e.g., "Matapeake, Unit xxx, Shelf xxx")

Source: Catalog of vibracores collected on Maryland's Continental Shelf
Example:

> "Vibracore collected by the U.S. Army Corps of Engineers as part of the Ocean City Beach Replenishment Project, Phase I (OCBRP-I) to locate & assess potential sand borrow areas for replenishment of the recreational beach at Ocean City, Md. Associated information: seismic data, grain size analysis"

SupplementalInformation

Definition: Information on how to access physical data represented by metadata record (e.g., general for entire collection, such as URL, or specific reference to online resource, like ordering system with specific ID) - mandatory
Value: "Contact the Program Chief of the Coastal & Environmental Geosciences Program at the Maryland Geological Survey, (410) 554-5500"
Source: n/a

Coordinates

Definition: Geographic coordinates (longitude, latitude), in decimal degrees – mandatory
Value: (-)decimal longitude, decimal latitude
Source: Excel spreadsheet, MMS_COREINFO.xls, fields "LongDD.DDD" and "LatDD.DDD"

AlternateGeometry

Definition: Alternate method of storing geospatial footprint; description of authoritative source of geographic location & how simple coordinates derived – optional
Value: none supplied
Source: n/a

OnlineResource

Definition: URL pointer(s) to textual information about specific record - optional
Value http://www.mgs.md.gov/coastal/osr/index.html
Source: n/a

BrowseGraphic

Definition: URL pointer(s) to images representing specific record - optional
Value: none supplied
Source: n/a

Date

Definition: Meaningful date (e.g., collection date) attached to record; may be to any degree of precision or left blank (e.g., 20010101, 1939-1945, -20030331, 2000-) - optional
Value: date of sample collection, from Excel spreadsheet, MMS_COREINFO.xls (Field "DateCollected"). In the spreadsheet, dates were reported in a variety of ways: as specific days, month and year, or range of months and years. Dates were converted to YYYYMM, YYYYMM-YYYYMM, or YYYYMMDD format by reordering records in ascending order by Collection Date, finding specific dates, one by one, in the original format, and replacing them

with dates in the new format. (Dates in the original format were not reported consistently and it was faster to find/replace than to figure out how to get Access to accomplish the task automatically.)
Source: Excel spreadsheet, MMS_COREINFO.xls, field "DateCollected"

DatasetReferenceDate

Definition: Reference date indicating currency of underlying data record (e.g., date metadata record added to National Catalog); format=YYYYMMDD - mandatory
Value: Date record provided to NGGDPP for uploading to National Catalog
Source: Provided by curator

VerticalExtent

Definition: Vertical extent (e.g., vertical depth information for rock core samples); contains 2-3 elements: unit of measure, max value, min value (e.g., m, 35.4, 0 => rock core measured at 35.4 total meters)
Value: Maximum value is the depth below the sediment/water interface to which the core penetrated; the minimum value, always zero, is the sediment/water interface
Source: Excel spreadsheet, MMS_COREINFO.xls, fields "CoreDepth" and "DepthUnit"
Example: "feet, 17.1, 0"

Location of Archived Sample

Continental shelf cores are stored on shelves at Matapeake, wrapped in sealed plastic sleeves, usually labeled with the researcher-assigned core ID and date of collection

References

Wells, D., and Conkwright, R., 1996, Physical inventory and repository of vibracores collected on Maryland's Continental Shelf: Baltimore, Md., Maryland Geological Survey, Coastal and Estuarine Geology Program File Report No. 96-6, 65 p.

APPENDIX 3
NGGDPP Metadata Fields:
Sediment Cores, Chesapeake Bay and Tidal Tributaries, Maryland
(Collection P1648)

(Specifically, the Chesapeake Bay Earth Science Study (CBESS) Cores)

Sources of Information:

- Comma-delimited ASCII file of lab results and derived statistics, Chesdata.asc, available for download from the MGS website (http://www.mgs.md.gov/coastal/data/baysedata.html)
- Lab notebook (hardcopy) used to track sample processing and post-processing storage
- Final report associated with the project (Kerhin and others, 1988)
- Field notebooks and associated maps (Oyster Bar Charts)

MetadataID

Definition: Metadata identification number
Value: 1 to 4255
Source: Assigned automatically by Microsoft Access; corresponds with field "ID" in data file Chesdata.asc.

CollectionID

Definition: NGGDPP collection identification number
Value: P1648 (Sediment Cores, Maryland Chesapeake Bay & Tributaries – P1648)
Source: DataPreservation.mdb – tblCollection – field "USGS ID"

Title

Definition: Official, human-readable title for individual record, used in listings & search results (short, distinctive) – mandatory
Value: CBESS, Grab #
Where
"CBESS" stands for the name of the project,
"Grab" indicates a short core, collected by a surface grab sampler, as opposed to a longer core, and
is the number assigned the sample in the data file Chesdata.asc. (field "ID")
Source: ASCII data file, Chesdata.asc

Alternate Title

Definition: Additional title identifiers for individual record (e.g., for further identification by other Web service interfaces); textual titles or specific sample IDs used by collection – optional
Value: A four-part sample identifier, separated by semi-colons; used by MGS at the time of data collection:

- "Chart #," Oyster Bar Chart number, from field notebooks and maps and the lab notebook
- "Raydist Red," from the data file Chesdata.asc. "Raydist Green," from the data file Chesdata.asc.
- "#," "Sample Number," from the data file Chesdata.asc. Note that "Sample Number"

does NOT uniquely identify a sample in the data set, nor does it correspond to the number assigned in the metadata field "Title." To create a unique sample identifier, it must be tied to the date of sample collection for mid-bay samples. For nearshore samples, it must be tied to Raydist Red and Green lane designations.

Explanation: In designing a sampling scheme, MGS researchers divided the Maryland part of the Chesapeake Bay into several dozen "tiles," based on a series of Oyster Bar Charts, numbered 0-36 (with gaps). Throughout processing and storage, samples remained grouped by those chart numbers. At the time of collection, a now-obsolete coordinate system, consisting of Raydist Red and Green lanes, was used to locate samples in the field. Upon collection, samples were labeled with the Oyster Bar Chart number and the date of collection and were then consecutively numbered, beginning anew each day with "1." (To further complicate matters, two sampling schemes were developed, one for shallower "nearshore" samples and the other for deeper water "mid-bay" samples, designations included in the metadata field "Abstract."

Sources: Field notebooks and maps, lab notebook, ASCII flat file, Chesdata.asc

Example: Chart 0; Raydist Red = 240.41; Raydist Green = 29.22; #42

Abstract

Definition: Human-readable description of individual record, used to help determine nature of underlying physical data resource; contains much information about data resource – mandatory

Value:

"Unprocessed, surficial sediment sample collected by the MGS as part of the Chesapeake Bay Earth Science Study (CBESS), a multidisciplinary research effort authorized by Maryland and Virginia through the U.S. Environmental Protection Agency (EPA). MGS collected samples throughout the Maryland Chesapeake Bay according to two sampling schemes, mid-bay and nearshore. This (mid-bay, nearshore) sample was processed to develop a baseline inventory of the physical and geochemical characteristics of bay bottom sediments. Since its collection, it has been stored in a plastic bag at ambient outdoor temperatures.

Associated information: grain size composition, water content, bulk density, carbon (total & organic) and sulfur content

Location: Matapeake, Unit xxx, Shelf xxx, Box"

SupplementalInformation

Definition: Information on how to access physical data represented by metadata record (e.g., general for entire collection, such as URL, or specific reference to online resource, like ordering system with specific ID) - mandatory

Value: "Contact the MGS Collections Curator, (410) 554-5500"

Source: n/a

Coordinates

Definition: Geographic coordinates (longitude, latitude), in decimal degrees – mandatory

Value: (-)decimal longitude, decimal latitude (derived from a concatenation of two fields in the data file Chesdata.asc. ("Decimal Longitude," and "Decimal Latitude").

Source: Data file Chesdata.asc.

AlternateGeometry

Definition: Alternate method of storing geospatial footprint; description of authoritative source of geographic location & how simple coordinates derived – optional
Value: "Coordinates are based on the North American Datum of 1927"
Source: n/a

OnlineResource

Definition: URL pointer(s) to textual information about specific record - optional
Value: http://www.mgs.md.gov/coastal/data/baysedata.html
Source: n/a

BrowseGraphic

Definition: URL pointer(s) to images representing specific record - optional
Value: none supplied
Source: n/a

Date

Definition: Meaningful date (e.g., collection date) attached to record; may be to any degree of precision or left blank (e.g., 20010101, 1939-1945, -20030331, 2000-) - optional
Value: date of sample collection, from ASCII data file, Chesdata.asc (Field "Collection Date"). Dates were converted from (M or MM/D or DD/YYYY format to YYYYMMDD format by reordering records in ascending order by Collection Date, finding specific dates, one by one, in the original format, and replacing them with dates in the new format. (Dates in the original format were not reported consistently and it was faster to find/replace than to figure out how to get Access to accomplish the task automatically.)
Source: ASCII data file, Chesdata.asc

DatasetReferenceDate

Definition: Reference date indicating currency of underlying data record (e.g., date metadata record added to National Catalog); format=YYYYMMDD - mandatory
Value: Date record provided to NGGDPP for uploading to National Catalog
Source: Provided by curator

VerticalExtent

Definition: Vertical extent (e.g., vertical depth information for rock core samples); contains 2-3 elements: unit of measure, max value, min value (e.g., m, 35.4, 0 => rock core measured at 35.4 total meters)
Value: "cm, 5, 0"
Source: Kerhin et al., 1988, p. 10

Location of Archived Sample

All remaining CBESS samples are stored on shelves at Matapeake, in plastic bags inside glass jars, inside cardboard boxes, labeled with an Oyster Bar Chart number, a mid-bay/nearshore designation, and the range of Raydist Red and Green lanes of the samples in the box

As of the date of this report, the curator must know the following to retrieve a particular sample:

- Type of sample (mid-bay or nearshore)
- Oyster Bar Chart number
- Raydist Red and Green coordinates
- Sampling date and sampling number

Samples were grouped, in order, by (1) sample collection technique (mid-bay/nearshore), (2) Oyster Bar Chart, (3) Raydist Green lane. For each Oyster Bar Chart, box numbering begins anew, with 1 (of x). Table # shows the box numbers and jar numbers within each box

Additional Information about the Sources, Samples, Etc.

Collecting the Samples

Mid-bay samples were collected on a 1-km grid and coarsely located by Oyster Bar Chart (Chart #). On any given sampling day, samples were assigned a Chart # and then numbered consecutively as they were collected, beginning with "1." Precise sample location was based on a now obsolete coordinate system, Raydist Red and Green lanes – the "R" and "G" designations on the labels. Raydist coordinates were later converted to LORAN-C, and, later still, to latitude and longitude.

Nearshore samples were collected in shallower water, along shore-normal transects at specified time intervals (based on constant boat speed) to a maximum water depth of ~10 ft. Those samples were also identified by Oyster Bar Chart (Chart #) and by the coordinates (Red and Green lanes) of the onshore starting point of the transect. Beach samples, if collected, were assigned a sample number of "0." Each successive sample taken along a transect was numbered consecutively, beginning with "1" for the underwater sample collected nearest shore.

References

Kerhin, R.T., Halka, J.P., Wells, D.V., Hennessee, E.L., Blakeslee, P.J., Zoltan, N., and Cuthbertson, R.H., 1988, The surficial sediments of Chesapeake Bay, Maryland: Physical characteristics and sediment budget, Baltimore, Md., Md. Geological Survey Report of Investigations No. 48, 82 p.

ASCII flat file, Chesdata.asc, available for download from the MGS website (along with associated metadata)

L. Hennessee, pers. comm.

Questions to Resolve

- Are lat/long reported in NAD27 or NAD83?

APPENDIX 4
MGS-Specific Instructions for Uploading Metadata to the National Catalog

(Refer to the NGGDPP document, *Preparing Metadata for the National Digital Catalog* (05/15/2009), for additional instructions.)

1. In the Microsoft Access Data Preservation database, copy tblMetadata_P###_etc to tblUploadMetadata_P###_etc

2. Change the names of the fields in tblUploadMetadata_P###_etc to correspond to the PROPERLY SPELLED field names expected by National Catalog; erase columns not included in the National Catalog.

   REMEMBER: In fields that allow multiple entries, like alternateTitle, entries should be separated by a COMMA. If the entries do not represent multiple, stand-alone titles, use a SEMI-COLON to separate them. For example, for the CBESS samples, MGS entered "Chart #, Raydist Red, Raydist Green, and Sample #," all of which taken together represent an alternate title for a sample. Initially, MGS used a comma to separate the four items. Consequently, the National Catalog recognized Chart # as one title, Raydist Red as another, etc. To rectify the problem, Rick Brown of the NGGDPP recommended using another character (e.g., a semi-colon) to separate the elements that make up a single alternate title.

   ALSO REMEMBER: In Access, if a form is used to populate a table, spacing lines using the ENTER key may interfere with file import into the National Catalog. (This remains to be tested.)

   It may be better to build a table with separate fields (e.g., Datum, Quad name) and then concatenate the fields in building the NGGDPP metadata table, rather than using a form to populate the table and leaving blank lines between items included in a single field.

   ALSO REMEMBER: If, during upload, the National Catalog returns an identical error message for each metadata record, check to be sure that the field name is spelled correctly (i.e., datasetReferenceDat**e** NOT datasetReferenceDat**a**)

3. Export tblUploadMetadata_P###_etc to .csv file format:
   a. Highlight tblUploadMetadata_P###_etc
   b. File => Export (select directory in which to store exported file and the Text Files type)
   c. In response to Export Text Wizard, choose:
      Delimited <Next>
      Delimiter = Other (|); Include Field Names on First Row; Text Qualifier = {none} <Next>
      Export to File: N:\NGGDPPGrant_2009\Metadata\Text\tbl tblUploadMetadata_P###_etc.csv
      (Change .txt to .csv) <Finish>

4. To upload the file to the National Catalog, go to http://my.usgs.gov/csc/nggdpp/upload, log in, and follow the instructions.

APPENDIX 5
MGS Data Preservation Advisory Panel
2009-2010 Membership
(√ denotes those in attendance at the inaugural Panel meeting on November 17, 2009)

√ **Mr. Tim Baker**, Deputy State Archivist
Maryland State Archives
Hall of Records
350 Rowe Blvd.
Annapolis, MD 21401
Office phone: 410-260-6402
Cell phone: 443-223-1685
Fax: 410-974-2585
E-mail: tbaker@mdsa.net

√ **Ms. Laura Bowne**, GIS Lead
MD Dept. of Natural Resources
Land Acquisition & Planning
Tawes State Office Bldg.
580 Taylor Ave., E4
Annapolis, MD 21401
Office phone: 410-260-8413
Cell phone: 410-279-3878
E-mail: LBowne@dnr.state.md.us

√ **Ms. Janet DeTore**, Project Geologist
Hillis-Carnes Engineering Associates, Inc.
10975 Guilford Rd., Suite A
Annapolis Junction, MD 20701
Office phone: 410-880-4788
Cell phone: 443-463-0031
Fax: 410-880-4098
E-mail: jdetore@hcea.com

√ **Mr. Eric Dougherty**, Chief
Engineering Geology Division
Office of Materials Technology
State Highway Admin.
7450 Traffic Dr.
Hanover, MD 21076
Office phone: 443-572-5171
Cell phone: 443-695-0939
E-mail: EDougherty@sha.state.md.us

√ **Ms. Carrie Gross**, Outreach & Appraisal Archivist
Maryland State Archives
Hall of Records
350 Rowe Blvd.
Annapolis, MD 21401
Office phone: 410-260-6464
Fax: 410-974-2585
E-mail: carolg@mdsa.net

√ **Mr. Jeffrey Halka**, Director
Maryland Geological Survey
2300 St. Paul St.
Baltimore, MD 21218
Office phone: 410-554-5503
Fax: 410-554-5502
E-mail: jhalka@dnr.state.md.us

√ **Ms. Lamere Hennessee**, Geologist
Maryland Geological Survey
2300 St. Paul St.
Baltimore, MD 21218
Office phone: 410-554-5519
Fax: 410-554-5502
E-mail: jhalka@dnr.state.md.us

√ **Ms. Maureen Kavanagh**, Chief
Office of Archeology/GIS
MD Dept. of Planning
MD Historical Trust
100 Community Pl.
Crownsville, MD 21032
Office phone: 410-514-7660
E-mail: mkavanagh@mdp.state.md.us

√ **Ms. Amanda Moore**, State Soil Scientist
U.S. Dept. of Agriculture
Natural Resources Conservation Service
John Hanson Business Center
339 Busch's Frontage Rd., Suite 301

Annapolis, MD 21409-5543
Office phone: 443-482-2913
Cell phone: 443-534-6358
Fax: 410-757-0687
E-mail: Amanda.Moore@md.usda.gov

√ **Mr. John R. Ninesteel**, GIS Specialist
Frostburg State Univ.
Tawes Hall Room 226
101 Braddock Rd.
Frostburg, MD 21532
Office phone: 301-687-4229
Fax: 301-687-3082
E-mail: jrninesteel@frostburg.edu

√ **Dr. James Reger**, Geologist (ret.)
2409 Deer Park Rd.
Finksburg,, MD 21048
Home phone: 410-861-8313
E-mail: jreger@dnr.state.md.us

√ **Mr. Dale Shelton**, Public Affairs Officer
Maryland Geological Survey
2300 St. Paul St.
Baltimore, MD 21218
Office phone: 410-554-5505
Fax: 410-554-5502
E-mail: dshelton@dnr.state.md.us

√ **Ms. Jennifer Stott**, Instructor
Howard Community College
Science & Technology Dept.
Room CL 145
Columbia, MD 21044-3197
Office phone: 410-772-4083
E-mail: jstott@howardcc.edu

√ **Mr. John H. Talley**, Director & State Geologist
Delaware Geological Survey
DGS Bldg.
Univ. of Delaware
Newark, DE 19716-7501
Office phone: 302-831-8258 (2833)
Fax: 302-831-3579
E-mail: waterman@udel.edu

√ **Ms. Ann Wheeler**, Librarian
MD Dept. of Natural Resources
Tawes State Office Bldg.
580 Taylor Ave., B-3
Annapolis, MD 21401
Office phone: 410-260-8830
E-mail: AWheeler@dnr.state.md.us

Ms. Marion Wiggins, Senior Geologist
Vulcan Materials Co.
875 Oxford Ave.
Hanover, PA 17331
Office phone: 717-637-7121 x498
Cell phone: 336-413-2676
Fax: 717-637-1234
E-mail: wigginsm@vmcmail.com

√ Mr. Peter Yencsik, Geologist
Minerals, Oil and Gas Division
MD Department of the Environment
1800 Washington Blvd., Suite 655
Baltimore, MD 21230
Office phone: 410-537-3557
Fax: 410-537-3573
E-mail: pyencsik@mde.state.md.us