US NATIONAL OCEANOGRAPHIC DATA CENTER ARCHIVAL MANAGEMENT PRACTICES AND THE OPEN ARCHIVAL INFORMATION SYSTEM REFERENCE MODEL

Donald W. Collins
US Department of Commerce/NOAA/NESDIS
National Oceanographic Data Center
1315 East West Highway, SSMC3 Fourth Floor
Silver Spring, MD 20910-3282
Donald.Collins@noaa.gov
+1-301-713-3272
+1-301-713-3302

Abstract
This paper describes relationships between the Open Archival Information System Reference Model (OAIS RM) and the archival practices of the NOAA National Oceanographic Data Center (NODC). The OAIS RM defines a thorough approach to defining the processes, entities, and framework for maintaining digital information in an electronic archival environment without defining how to implement the framework. The NODC Archival Management System (AMS) is an example of an implementation of a persistent digital archive. Major OAIS RM components, such as the Submission Information Package, Archival Information Package, Dissemination Information Package, and Archival Storage are clearly comparable between the OAIS RM and the NODC AMS. The main participants (Producer, Consumer, Management, and OAIS) are represented in the NODC AMS, as are many primary functions (Ingest Process, Archive Process, Dissemination Process). Some important OAIS RM components, such as a consistent Submission Agreement and a deeper level of Preservation Description Information may be missing for some of the information archived in the NODC AMS. It is instructive to document the commonalities between the NODC system and the OAIS RM as the NOAA National Data Centers expand archival services for a broad and growing range of digital environmental data.

1 Introduction

Imagine that you are on a large ship on a calm sea with no land in sight and 3 miles of water beneath your feet. You are a marine biologist studying the habitat of the giant squid, so you want to know the range of water temperatures and salinity values, the available nutrients eaten by local microscopic organisms, what those organisms are and what visible-size sea life might be eating them. These pieces of information can be collected by a host of instruments you deploy over the side of the ship and lower through the water. The ship stops periodically to collect these measurements, which are electronically recorded into a series of data files that you will use later in the lab to characterize the conditions in which the giant squid lives.

If you received funding for your research from the US Federal government, you are likely to be contractually obligated to send a copy of the data you collected to the National Oceanographic Data Center (NODC), along with enough descriptive metadata to make the data meaningful to others. The NODC is one of several data centers operated by the

U.S. Department of Commerce National Oceanic and Atmospheric Administration (NOAA).

The NODC is believed to archive the largest collection of *in situ* measurements of oceanographic parameters in the world, with approximately 300 gigabytes of data, metadata, and model output stored in its digital archives (D. Knoll, NODC, pers. comm., 2003). The NODC has recently undertaken a substantial effort to update and improve the archival management of these electronic environmental records and to improve access to the original data records in its collection. This paper examines how these new record management strategies and procedures relate to the Open Archival Information System Reference Model (OAIS RM) [1], ISO 14721:2002, which defines the major elements and functions of an electronic records archive. To explore the relationship between the OAIS RM and the new NODC Archives Management System (AMS), the first part of this paper provides an overview of the main features of the OAIS RM. The second part of the paper describes in some detail the NODC AMS in terms of the conceptual elements of the OAIS Reference Model. The last part of the paper discusses OAIS RM elements that are not yet implemented by the NODC AMS.

2 OAIS Reference Model

The OAIS Reference Model was initially developed by the Consultative Committee for Space Data Systems to identify the essential elements needed by an electronic records archive to manage records over long time scales [2]. The OAIS RM acknowledges technological changes that inevitably create substantial challenges for electronic records archives [3]. As described by Lavoie [4], an OAIS means "any organization or system charged with the task of preserving information over the Long Term and making it accessible to a specified class of users (known as the Designated Community)." In keeping with the practice established in the OAIS RM documentation, entities and procedures named in the model and discussed below will have initial capital letters (e.g., Producer). By definition in the OAIS RM (p. 1-11), data are archived for the Long Term, which is defined as "[a] period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future."

The main participants in any OAIS are defined as Producer, Consumer, Management, and the OAIS archive functions. In the OAIS RM, Producer represents the people or systems that create and/or provide information to be preserved in the archive. Consumer represents the people or systems that use the OAIS to find and retrieve preserved information. A Designated Community may be a special subset of the domain of all Consumers. Management is the person or group that sets the policy for the management of the OAIS, among other activities: it is not responsible for the day-to-day operations of the OAIS. The OAIS (archive) is the system that provides for the long term storage, migration, and dissemination of the information that is archived. Using this model and the example in the Introduction to this paper, the following roles could be identified:

- Producer: Ocean scientists (e.g., squid specialist) making measurements of physical, chemical, and biological conditions;

- Consumer (Designated community): other ocean scientists, resource managers, general public;
- Management: US National Oceanographic Data Center Director and Deputy Director;
- OAIS (archive): US National Oceanographic Data Center Archive Management System.

One of the basic concepts of the Reference Model is that information is a combination of data and its Representation Information. Regardless of whether the data part of the information is physical (e.g., a giant squid) or digital (e.g., digital photographs of the giant squid), Representation Information allows the Consumer to fully interpret the meaning of the information. For digital objects, Representation Information typically includes some mapping of bits into recognizable data types (e.g., characters, integers). It also associates these mappings into higher-level groupings of data types, which are called Structure Information. To fully understand how to interpret the Structure Information, it is important to include Semantic Information, which defines the language of the Structure Information [5].

Another main component of the OAIS RM is the Information Object, which is comprised of the following components:

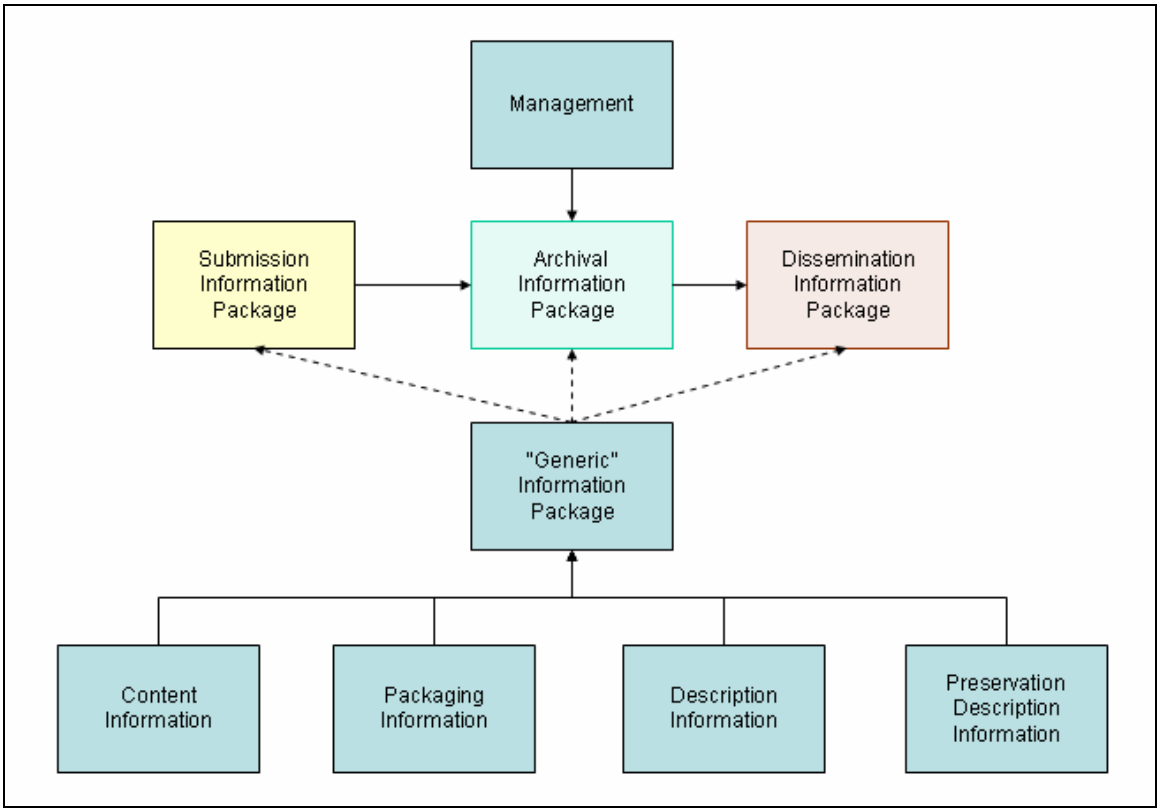
- Content Information object, which is equivalent to the contents of a letter, email, observed data values, etc.;
- Packaging Information object, which is information that wraps the different information objects into a cohesive bundle;
- Descriptive Information object, which provides descriptive content and contextual information about the Content Information object;
- Preservation Description Information object, which includes significant information about the electronic form and structure of the bits that can be translated into the Content Information object. It also includes information about provenance and authenticity validation characteristics of the Content Information object.

The OAIS RM distinguishes between Information Objects based on the role that the Information Object performs in the information management process. Three specific types of Information Object are designated: the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP). Each Information Package contains the four information objects defined above. Figure 1 depicts the relationship between the components of an Information Package and between SIP, AIP, DIP, and Management (after Sawyer [6]).

A Producer creates electronic records: the contents are the property of the Producer and may be in any format that is deemed useful by the Producer. Once the Producer decides to transfer the data to the OAIS, a Submission Agreement is negotiated between Producer and OAIS to define the terms of the information transfer. The OAIS may help the Producer in developing the SIP by providing information, tools, or other assistance in preparing the contents of the SIP, especially the Preservation Description Information

component. The SIP contains Content Information, but must also contain sufficient metadata to ensure that the Content Information can by maintained properly by the OAIS and be used by future Consumers [7]. The SIP is then transferred to the OAIS in one or more Data Submission Sessions.

Fig. 1. OAIS Reference Model Information Package components (After Sawyer, 2002.)



Upon receipt of the SIP, the OAIS creates an Archival Information Package (AIP) using the archives' ingest procedures. The OAIS defines the appropriate ingest procedures for creating the AIP according to the archives policies and guidelines. The OAIS may modify the form and content of the SIP: "An OAIS is not always required to retain the information submitted to it in precisely the same format as in the SIP. Indeed, preserving the original information exactly as submitted may not be desirable [8]". The intention here is to ensure the preservation of digital information, not to modify or tamper with the digital content. Once the SIP is transformed into the corresponding AIP and Package Descriptions (i.e., the information needed to make an AIP accessible from appropriate Access Aids) during the ingest process, the AIP is stored in an Archival Storage entity. Ingest processes, Package Descriptions, and Archival Storage hardware and software may vary significantly from one OAIS to another.

The OAIS determines how to make the AIPs in its collection available to its Designated Community. To perform this function, OAIS defines a Dissemination Information

Package (DIP). The DIP is described in the Package Description, found using Access Aids provided by the OAIS. Access Aids are the tools provided to discover and obtain AIPs from the OAIS. Finding Aids, Collection Descriptions, Ordering Aids and other data discovery tools are types of Access Aids. When a user discovers the existence of archived materials through available Access Aids, the selected AIPs are assembled into a Dissemination Information Package (DIP) and transferred to the Consumer via a Data Dissemination Session. The structure and mechanism for delivering a DIP to Consumers depends on the way the archival organization creates its DIPs [9].

This introduction to the main components of an OAIS only scratches the surface of an extremely detailed and well-defined set of terms, objects, concepts, and procedures that an electronic archives needs to address to ensure the preservation of data and information for an indeterminate "Long Term". The next section of this paper describes the archival practices established by the US NODC and relates many of the functions, processes and information objects of the new NODC Archives Management System to the components of the OAIS Reference Model.

3 NODC Archives Management System: A Case Study

The US National Oceanographic Data Center (NODC) was created as an inter-Departmental support organization administered by the US Navy in 1960 and transferred to the Environmental Science Services Administration (ESSA) in the mid-1960s. When the National Oceanic and Atmospheric Administration (NOAA) was created in 1970, the NODC became one of the three environmental data centers administered by the NOAA. The NODC receives oceanographic data from a diverse community of international oceanographic organizations, government organizations (federal, state, and local), and public and private universities and research institutions from around the world.

The primary commonality among these organizations and the data they provide to the NODC archives is that the information is somehow related to the world's oceans, seas, and coastal areas. Data formats and structures, languages used, and the types of data collected can be extremely variable. Most oceanographic data collected in the past decade or two are provided to NODC in digital data files. As technology changes, new equipment for obtaining, storing and organizing measurements are invented. Data formats and structures change to accommodate new measurable values and techniques. The only constant throughout these technological changes is the need for accurate metadata about the instruments used (e.g., calibration and methodologies), data format structure and other documentation.

By 1967, the NODC recognized the need to improve tracking of data that were sent to the NODC. A data set identification system was developed in which groups of data were assigned an NODC Accession Number to identify the information as a unit. An accession is loosely defined as 'a logical grouping of related data,' which is usually interpreted as a group of data that are received together. Typical examples of the types of data in an accession include: *in situ* water column measurements (e.g., water temperature and salinity, nutrient concentrations such as dissolved nitrate and silicate, current speed and

direction), biological observations (e.g., abundance and taxonomic identification of plankton and fish species), or satellite observations of sea surface characteristics.

As of December 31, 2003, there were 20,419 individual accessions in the NODC archives. New accessions are presently acquired at a rate of about 30 per month. During a recent month, more than 62 Gigabytes of data were downloaded by 1088 individual host computers connected to online services that access NODC information, holdings and products created from those holdings (NODC Information Systems and Management Division, unpublished report).

The NODC recently developed the NODC Archives Management System (AMS) to bring the electronic and analog data, metadata, and administrative information files for oceanographic data collections into a more robust and flexible environment. The Accession Tracking Data Base (ATDB), the Archive File Management System (AFMS), the Ocean Archive System and the NODC Metadata Repository (NMR) are the primary components of the Archives Management System. The NODC Metadata Repository is a commercially-available database (using a proprietary structure and database developed by Blue Angel Technologies, Inc. and Oracle Corp.) that is optimized for managing data set descriptions in the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) structure. While this component will soon provide descriptive metadata to assist with search and retrieval processes, the underlying software and database of this component of the AMS are beyond the scope of this paper. The emphasis in the next few sections is to describe the AMS components designed, developed and deployed within NODC on generic computer hardware using open source software for the operating system, database design and management, and preliminary data entry, search, and retrieval requirements.

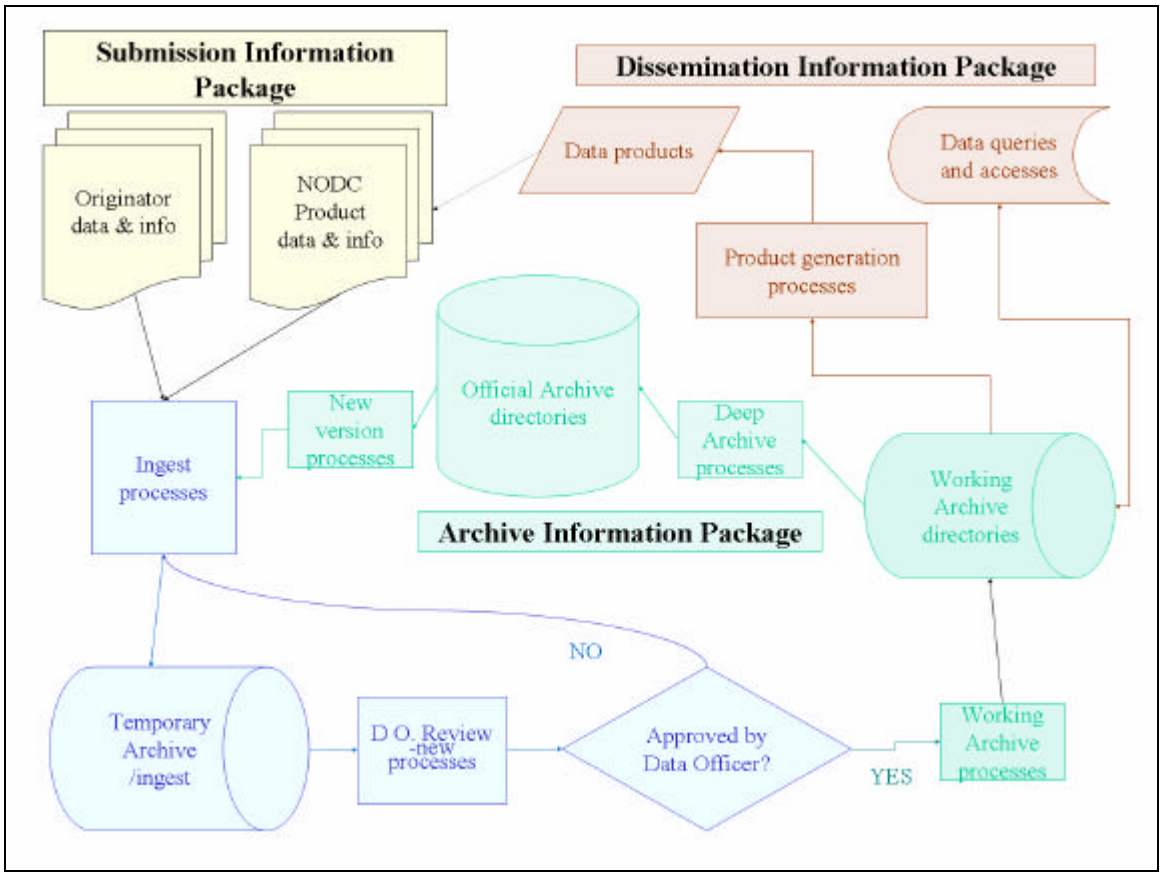
The NODC AMS identifies a series of information elements, procedures, and standardized practices that facilitate ingesting, describing, accessioning, storing, migrating, and accessing archived digital records and information. These functions are discussed below and are related to relevant OAIS RM concepts or constructs. Figure 2 depicts the generic flow of an Information Package through the Archives Management System.

3.1 Ingest Procedures

The NODC frequently works with data Producers to prepare their information for archiving, i.e., to create a more meaningful Submission Information Package. At the NODC, the SIP is equivalent to a single accession as it is received from the Producer (internally, NODC currently refers to "originators data" rather than "a SIP"). During this interaction, the NODC encourages the data Producer to provide the most generic representation of their data possible and to provide as much descriptive documentation as possible. The reason for this is to obviate the need for NODC to undertake after-the-fact translation tasks to represent data that are in software and/or platform dependent data structures into less dependent structures. In many instances, these translations can be undertaken with little or no loss of meaning for the data. The NODC prefers to have the

data Producer make these translations whenever possible, rather than making such translations during the ingest process.

Fig. 2. NODC AMS flow diagram for an Information Package.



As shown in Figure 2, a Producer collects data and prepares it in some fashion for shipment to the NODC. Of course, the method of "shipment" and the medium used in this Data Submission Session has changed substantially over time from mailing analog records, computer punch cards, or magnetic tapes to FTP transfers, web site downloads, or sending CD-ROMs or DVDs. Ingest procedures now in place for each Data Submission Session are:

- obtain and review the data/information files and/or other information objects in the SIP from the Producer;
- create a new record in the NODC ATDB (described below), which establishes the 'canonical structure' for the accessioned SIP in a preparation area;
- copy/move digital data files to the appropriate directory of the Archives File Management System (described below);
- request that the ATDB record be closed;
- review of 'closed' accession record and SIP by 'Data Officer';
- transfer approved SIP to the archival storage area (creation of the AIP).

Once a SIP is ingested and accessioned and an AIP is created, the data in it may be processed into an NODC product (e.g., the NODC World Ocean Database, which contains more than 7 million temperature, salinity, and other parameter profiles) or delivered as a direct copy (DIP) using the NODC Ocean Archive System, an online discovery and retrieval service (OAIS Access Aid) described in greater detail below.

3.2 Accession Tracking Data Base

The Accession Tracking Data Base (ATDB) is a relational database that supports initial SIP ingest procedures and administrative metadata management. The ATDB uses postgreSQL database management software on a generic server platform running the Linux operating system. Generically-designed browser-based user interfaces were developed using perl/CGI scripts to facilitate data entry and limited search/retrieval capabilities. Linux, postgreSQL, and perl are all freely available open source software designed to operate on a number of hardware platforms. One of the guiding principles for developing the ATDB and AMS was to determine if open source software tools were robust enough to support a mission critical information management system, rather than relying on commercial packages that require continuous license upkeep, maintenance and other expenses. The initial results suggest that the open source tools are more than adequate for the purposes of creating a workable file management system and a database to manage thousands of digital files. The ATDB represents several components of the Administration function in the OAIS RM [10].

A 'Brief Access Record' (BAR) is created for each new SIP using the main table in the ATDB. The BAR allows a new entry to be made in the ATDB based on a relatively few administrative and descriptive metadata elements. Administrative metadata elements keep track of an accession in the NODC File Management System and other pieces of internal information of importance to NODC but that may not necessarily be useful for Consumers. Administrative metadata, combined with descriptive metadata elements (discussed later) are roughly equivalent to the Package Description in the OAIS RM. Administrative metadata elements in the ATDB are:

- accession number (unique identifier for an Archival Information Package, automatically generated by the database),
- date received (date the SIP was received at the NODC),
- keyer, editor (name of the NODC employee creating the ATDB entry and making the most recent change to the ATDB entry, selected from a controlled vocabulary)
- keydate, editdate (system-assigned date-time stamp of ATDB entry creation and latest ATDB entry change),
- status (information indicator that denotes the 'state' of the Archival Information Package, i.e., 'new', 'archived', 'revision'),
- version (information indicator that denotes the most recent version of the AIP),
- availability date (date after which this AIP can be made accessible to the public; default value is the same as the 'date received'),
- requested action (information indicator to request action from the Data Officer function, i.e., 'close' ('assess this SIP and metadata for inclusion in the archives') or 'open' ('check the AIP out of the archive for revision, creating a new version of the AIP'),

- NODC contact (name of the NODC employee who is most familiar with the contents of this SIP or AIP. May be different than the 'keyer' or 'editor').

The ATDB also requires the inclusion of a minimal amount of descriptive metadata about the SIP. Available descriptive information is passed from the ATDB via a program to the NODC Metadata Repository, creating a partially-complete FGDC CSDGM-structured description of this AIP. Descriptive metadata elements in the ATDB include a title, start and end dates, latitude and longitude bounding coordinates, and controlled-vocabulary descriptors for institution names, sea areas, parameter and observation types, instrument types, project names, and platform (ship) names.

About 13 staff members are designated as keyers which means that they can create the initial BAR in the ATDB, although the majority of new BARs are created (at present) by 3 or 4 people. Four senior, non-management staff members are designated Data Officers. The Data Officers are responsible for reviewing the work of the keyers, approving new accessions for inclusion in the archives, and developing related archival management policy recommendations for approval by NODC Management.

3.3 NODC Archives File Management System

Creating a new record in the ATDB causes a perl script to create a directory structure on a storage disk (represented in Figure 2 as "Temporary Archive" and referred to internally as the "ingest area" or "/ingest") in the canonical directory structure (Table 1). While still in the ingest area, it is possible to modify or add to the contents of a SIP. Once the contents of the SIP are finalized, a program is run that copies files from the ingest area to the "archive area" or "/archive". This Archival Storage is presently provided by a large capacity RAID device with limited write privileges, i.e., only the archivist function (in most cases, via a program) can write files to this system. The canonical structure of an NODC AIP is represented in Table 1.

Table 1. NODC AMS File Management System canonical structure. Elements listed in *italics* are only part of the final archival version of the file structure. A "/" indicates the listed element is a directory.

7-digit unique number/	NODC Accession number (e.g., 0000001)
0000001.01-version.md5	*File containing checksum values for all files in this AIP.*
1-version/	*Directory identifying the most recent version of this AIP, beginning with 1-version.*
NODC_ReadMe.txt	Text file that describes this directory structure.
about/	Directory for storing information files created by NODC about this AIP.
journal.txt	NODC-created file describing actions taken by NODC staff regarding this AIP.
other files…	Optional other files created and/or maintained that are not part of the data in this AIP.
data/	Directory for storing the original data and translations of the original data in the AIP.

0-data/	Directory for storing an exact copy of the original files in this AIP.
directories or files	Copy of the original files obtained from the Producer.
1-data/	Directory for storing translated versions of the original files in this AIP.
directories or files	Translated versions of the original files in this AIP in a directory structure that may resemble the original directory structure.

The main sections of the canonical AIP structure are /about and /data. Any additional information about the SIP, such as emails between NODC and Producer or other NODC-created information about the SIP, are placed in the /about subdirectory. The journal.txt file, which is initially created along with the directory structure, is used to document any steps taken by NODC personnel while performing ingest processes on the SIP. At present, NODC guidelines require that an exact copy of the original files in a SIP are copied to the /0-data directory, regardless of their original file structure. The /1-data directory is where NODC-created translations of the original SIP contents may be placed. The intention of this directory is to provide a place for non-proprietary representations of proprietary original data from the SIP. A typical example of the contents of /1-data might be the comma-separated ASCII representation of the text and values from a Microsoft Excel spreadsheet file. In an effort to minimize the probable lack of access to this file due to program changes or loss of vendor support for the program in the future, NODC attempts to translate files from proprietary structures to generic structures.

3.4 Archiving Procedures

The NODC continues to refine its archiving procedures for digital accessions. Archiving processes are undertaken by the employees designated as Data Officer. As noted above, Data Officer tasks (see Figure 2) include reviewing the work of the keyers and determining if an accession is ready to be moved from the /ingest area to the /archive area. There are currently only four criteria that the Data Officer uses to make this determination:

- there is a fully-populated ATDB entry,
- the journal.txt file has been updated,
- the files in the /ingest directory are in the canonical form (Table 1),
- a "reasonable attempt" has been made to translate data from proprietary formats to generic formats and translations are placed in /1-data.

When a keyer determines that the ingest-AIP needs no further action, the ATDB element ‘requested action’ is set to ‘Close’. This indicates to the Data Officer that the SIP is ready for review. If the four criteria listed above are met, the Data Officer approves the transfer of the SIP from the ingest area to the Working Archive Directories (Archival Storage), creating the 'working archive' copy of the AIP (Figure 2). The program that transfers files from /ingest to /archive calculates a checksum value for each file, which can be used for future validation of the contents of the AIP, and also runs additional virus-detection software to minimize the chance of archiving a virus with the AIP.

As noted in Figure 2, there are also "Deep Archive processes" that include the creation and validation of off-site copies of the AIP (one is maintained in Asheville, NC at the National Climatic Data Center and one will soon be maintained at the NODC National Coastal Data Development Center in Stennis Space Center, MS). These 'deep archive' copies are intended for use in disaster recovery situations or when the local 'working archive' copy is rendered temporarily unavailable due to equipment malfunction or other reasons. This backup process represents the Replication function described in the OAIS RM [11].

Figure 2 also shows a 'New Version processes' step. These processes, which require approval from the Data Officer, are used when a new version of an existing AIP is required. New versions are occasionally requested by a Producer, usually when an error is found in a previously submitted data set. In cases where a new version is requested, an exact copy of the existing AIP is 'checked out' of the archive area and placed back in the ingest area. Modifications are then made to the data files and/or metadata files by the editor who requested the new version. When all modifications have been made, the same approval process is followed by the Data Officer, with special attention paid to the documentation of what modifications were made, why they were made, and who made them. The entire AIP is then 'checked-in' to the Working Archive Directories as "/2-version" (or the next available version number). In this fashion, NODC maintains a copy of each iteration of a specific data set. Circumstances for determining if an update SIP should be a new version of an existing AIP or should be a new AIP are decided on a case-by-case basis by the Data Officer and Management. The current philosophy adopted by NODC management is that it is better to keep everything, including obsolete versions, than to be unable to provide back to the originator an exact copy of everything that was given to the Center.

3.5 Metadata Management

Descriptive metadata for new AIPs, in the form of FGDC CSDGM-compliant database records, are presently created and/or maintained in two places: the ATDB Brief Access Record and the NODC Metadata Repository (NMR). The BAR is discussed in detail above. The NMR is beyond the scope of this paper, but will be used to manage additional descriptive metadata for each AIP in the NODC collection. There is some overlap between the information in each of these databases. NODC maintains several authority tables (e.g., people, ship names, institutions, place names) in the ATDB to facilitate the creation of consistent descriptions of an AIP. Controlled vocabulary entries will only be updated through the ATDB. Details of how to propagate any updates efficiently into the NMR are still in development. Ultimately, NODC plans to maintain a comprehensive accession tracking and descriptive metadata database from which any number of formatted descriptive information (e.g., FGDC, ISO 19115, Dublin Core) could be exported. As mentioned above, descriptive metadata is part of the OAIS concept of Package Description.

The OAIS RM requires a substantial level of semantic and syntactic metadata to be maintained in the Preservation Description Information (PDI) object of the AIP. At present, this level of metadata and the existence of a PDI object in an AIP depend on the

inclusion of such information in the SIP from the Producer. NODC is investigating how to create or refine the contents of this critical information element.

3.6 Data Dissemination

The NODC distributes two general types of Dissemination Information Package (DIP). One type of DIP is an exact copy of an AIP, with no additional processing beyond the possible translations described above. The other type of DIP distributed by the NODC is a Derived AIP, typically data from multiple AIPs processed by a product developer (usually within NODC, but occasionally by an external organization) to create a value-added data set, referred to as an NODC Standard Product. In most cases, all of the data in an NODC Standard Product have been reformatted to a single data format and possibly had some type of quality control checks performed, such as marking values that erroneously appear in a land area or a measured value that is outside the range of possible values for a parameter. The obvious advantage of these products is that similar types of data that were originally in a variety of data structures may now be easily inter-compared or otherwise statistically manipulated.

The OAIS RM categorizes Access Aids as a Finding Aid or an Ordering Aid. A Finding Aid allows a consumer to search for and discover DIPs that are of interest to the consumer. Ordering Aids are applications that help the Consumer to obtain DIPs of interest and include information costs and other handling circumstances that may be needed to transfer a copy of the DIP to the consumer. Standard products from the NODC (primarily Derived AIPs that are copied to CD-ROM or DVD and mass-replicated) can be discovered through a number of Finding Aid-like applications (e.g., NOAAServer, an FGDC Clearinghouse application), ordered from the NOAA National Data Centers Online Store [12], or discovered and downloaded using the NODC Ocean Archive System (OAS) [13].

The OAS is the current interface available directly from the NODC for searching, discovering, and retrieving copies of original data from the AMS. The OAS interface presents several of the ATDB descriptive metadata elements in a tabular form and allows a Consumer to select one or more descriptive elements to create a database query. Once the query has been processed, a list of accession numbers and descriptors for AIPs that match the query are presented, in addition to links to the full ATDB record and to the File Management System working archive directory. The Consumer can assess the relevance of the contents of the AIP and save all or part of the AIP to a local directory.

A small reference team is available to assist a Consumer in deciding whether there is a Standard Product that is more suitable for their needs or to assist with finding appropriate AIPs. In general, the majority of requests for data and information are satisfied by a Standard Product. However, as AIPs become readily available online DIPs through the OAS or other Access Aids, the demand for additional assistance for Consumers in deciding about appropriate data is expected to increase. In general, the Consumers for data from the NODC are: ocean scientists of all types from all over the world and all levels of government (the Designated Community), non-scientist business persons (often

lawyers and insurance representatives) and the general public (usually for K-12 educational purposes or recreational purposes).

4 What is missing?

The NODC Archives Management System described above is still in development, although the ATDB and File Management System parts of the AMS have been used operationally since about April 2002. The Ocean Archives System Finding Aid was released for public use in December 2003. Additional descriptive metadata elements were added to the ATDB in mid-2003 to accommodate metadata from a legacy database. But as noted above, the links between the ATDB and the NODC Metadata Repository (NMR) are not yet fully developed, although the same information used to load the ATDB with legacy information was used to populate the NMR for each AIP. Many issues related to maintaining the referential integrity of the information that resides in both the ATDB and the NMR are still being examined.

The discussion above outlines how the main elements of the NODC Archives Management System map to many of the major components of the OAIS Reference Model. However, some important OAIS RM components are missing from the NODC AMS. In particular, two areas that need additional work are the development of a Submission Agreement and identifying Detailed Preservation Information for each AIP in the AMS.

In most cases, no officially-communicated Submission Agreement spells out the terms, conditions, and other responsibilities of the NODC to act as the Long Term custodian of an Information Package. The NODC negotiates and maintains data and information exchange agreements with a number of US and foreign organizations as a routine part of its archives efforts, but many data sets are submitted to the NODC with little more than an email from the Producer or designated intermediary and varying levels of descriptive metadata. It is unclear if these sometimes informal cover letters are sufficient to serve as an OAIS Submission Agreement. The NODC participates in the development of the NOAA Comprehensive Large Array-data Stewardship System (CLASS), which is developing a Submission Agreement on behalf of the NOAA National Data Centers [14]. The draft version of the CLASS Submission Agreement is not yet available for public review, but currently is modeled using the FGDC Content Standard for Digital Geospatial Metadata (CSDGM) as a framework for defining a variety of custodial descriptive metadata. It is not yet clear if a separate formal Submission Agreement will be developed to authorize transferring data and information from the Producer to the NODC in addition to the current draft Submission Agreement.

Perhaps the most difficult missing element in the NODC AMS is the availability of detailed Preservation Description Information. The OAIS RM defines this as "[t]he information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context Information [15]". Some of this missing information will be provided for future accessions if something like the draft CLASS Data Submission form accompanies each Submission Information Package, but what about the historic data that already reside in the NODC

collections? Approximately 2700 of the 20,000+ AIPs archived at NODC have some type of information about their provenance and context in a Data Documentation Form (DDF), which was a standard form used to document data submitted to the NODC from the early 1970s until the mid-1990s. The quantity and quality of the information in the DDFs varies significantly and is dependent on the effort made by the Producer to document the data; nearly all existing DDFs are analog documents stored either on-site at NODC or in an off-site storage facility. In OAIS terms, most historic AIPs at NODC have little or no level of Fixity Information (which "…authenticates that the Content Information has not been altered in an undocumented manner [16]"). Likewise, there are very few instances where Representation Information (i.e., "information that maps a Data Object into more meaningful concepts... [such as] the ASCII definition that describes how a sequence of bits is mapped into a symbol [17]") is present at all. The NODC is beginning to consider how to address these difficult information deficiencies as part of the planning to maintain its collections of data for the Long Term.

5 Conclusions

The Open Archival Information Systems Reference Model describes a very thorough approach to defining the processes, entities, and framework for maintaining digital information in a electronic archival environment without defining how to implement the framework. The NODC Archival Management System provides an example of how to implement a persistent digital archive for oceanographic data. Many of the processes and components correlate well with elements of the OAIS Reference Model. Major components, such as the Submission Information Package, Archival Information Package, Dissemination Information Package, and Archival Storage are clearly comparable between the OAIS RM and the NODC AMS. The main participants (Producer, Consumer, Management, and OAIS) are also all present in the NODC AMS, as are many of the primary functions (Ingest Process, Archive Process, Dissemination Process). On the other hand, the NODC AMS is frequently lacking some important OAIS RM components, such as a consistent Submission Agreement and a deeper level of Preservation Description Information.

This paper establishes a relationship between the OAIS Reference Model and the archival management practices of the NOAA National Data Centers. It is important to document the commonalities in the NODC system and the OAIS RM as NOAA and the NOAA National Data Centers continue to develop and upgrade archival services for a broad and growing range of digital environmental data. While there are many commonalities between the NODC AMS and the OAIS RM, the NODC (and by extension, the NOAA National Data Centers) needs to be aware of the types of information that are not presently available or actively acquired for data that are sent to be archived. It is imperative that these environmental data records not become the "write once, read never" records bemoaned by Barkstrom [18] because they will provide the baseline scientific data for future environmental investigations and future generations.

6 References Cited

[1] Consultative Committee for Space Data Systems, 2002, Reference Model for an Open Archival Information System OAIS. Available online at

http://wwwclassic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf (last accessed April 2003).
[2] Lavoie, Brian, 2000, Meeting the challenges of digital preservation: The OAIS reference model. OCLC Newsletter, No. 243 (January/February 2000), p. 26-30. Available online at http://www.oclc.org/research/publications/newsletter/repubs/lavoie243/ (last accessed April 2003).
[3] Consultative Committee for Space Data Systems, 2002, p. 2-1.
[4] Lavoie, Brian, 2000, p. 27.
[5] Consultative Committee for Space Data Systems, 2002, p. 4-21.
[6] Sawyer, Donald, 2002, ISO "Reference Model for an Open Archival Information System (OAIS)": Tutorial Presentation. Presentation to University of Maryland College of Information Studies, October 2002, 31p.
[7] Consultative Committee for Space Data Systems, 2002, p. 4-49.
[8] Consultative Committee for Space Data Systems, 2002, p. 4-49.
[9] Consultative Committee for Space Data Systems, 2002, p. 4-52.
[10] CCSDS, 2002, p. 1-7 and p. 4-10.
[11] CCSDS, 2002, p. 5-5.
[12] NOAAServer is available online at http://www.esdim.noaa.gov/noaaserver-bin/NOAAServer (last accessed January 2004). The NOAA National Data Centers Online Store is available online at http://www.nndc.noaa.gov/dev/prototype/nndcserver/nndchome.html (last accessed January 2004).
[13] NODC Ocean Archive System can be accessed at http://www.nodc.noaa.gov/search/prod/ (last accessed January 2004).
[14] Habermann, Ted, (in prep.), Comprehensive Large Array-data Stewardship System (CLASS) Data Product Submission Agreements, 20p.
[15] CCSDS, 2002, p. 1-12.
[16] CCSDS, 2002, p. 1-10.
[17] CCSDS, 2002, p. 1-13.
[18] Barkstrom, Bruce R., 1998, Digital archive issues from the perspective of an Earth science data producer. NASA Technical Report, NASA Langley Research Center Atmospheric Sciences Division, Available online from http://techreports.larc.nasa.gov/ltrs/papers/NASA-98-dadw-brb/ (last accessed April 2003).

Acknowledgement
The author would like to thank Lauren Brown and the students in the “Seminar in Archives, Records, and Information Management” course at the University of Maryland College of Information Studies for their constructive comments. Also, many thanks to Kurt Schnebele, Tony Picciolo, Steve Rutz, Mary Lou Cumberpatch, Anna Fiolek, Bob Gelfeld and Donna Collins for their support, constructive observations and suggestions.