

Mapping, Cross-walking, Converting and Exchanging Oceanographic Metadata Information in Video Data Management System

Anna Fiolek, NOAA Central Library, Silver Spring, MD 20910

Susan Gottfried, Stacy Ladnier, Sharon Mesick,

NOAA National Coastal Data Development Center, Stennis Space Center, MS 39529

Abstract- The NOAA Central Library (NCL) has been collaborating with the Office of Ocean Exploration and Research (OER), and two divisions of the National Oceanographic Data Center (NODC) - the Marine Data Stewardship Division (MDS) and the National Coastal Data Development Center (NCDDC) - to address the requirements for metadata usage, archiving, preserving, managing, and providing online access to digital videos and still images from OER oceanographic expeditions.

The Video Data Management System (VDMS) was developed to create and manage the metadata information to provide online and off-line access to video and images obtained during signature (NOAA-sponsored) oceanographic expeditions. NOAA scientists can discover thousands of digital videos and still images, hundreds of digital video clips, highlight movies, and related documents and products via metadata included in NOAAALINC, the NCL online catalog (<http://www.lib.noaa.gov/>).

During the development of the VDMS project plan, which is part of a larger comprehensive OER Data Management Project, the VDMS team defined and established several 'best practices' to support OER video data management and metadata requirements. Metadata guidelines developed for digital video (DV12) and digital still images (DI12) help scientists and data managers in the field to create complete descriptive and technical metadata about their video and image data. Scientists, librarians and archivists then use this information to create the Federal Geographic Data Committee (FGDC), MARCXML or MARC21 metadata records. The VDMS team also developed a work-flow for creating and managing metadata information on digital video by defining the process for moving video data from ship to library to archive, including steps for creating

archival backup copies and web-accessible video clips and highlights.

Many types of metadata are required to ensure that digital video and still images can be interpreted and used by software that may be available on future computers and operating systems. Technical metadata about the encoding format, codices used, playback rates, etc., are critical for software to be able to correctly interpret and display the contents of the digital files. Descriptive metadata is necessary to provide context for the content being displayed and answers questions about when, where, why, and by whom the video and images were obtained. It is often easiest to acquire these types of descriptive and technical metadata from the principal investigator (PI) or data manager soon after each cruise.

This paper discusses a process of mapping different metadata schemas, metadata cross-walking and exchange between the NOAA Central Library Network, NODC, and NCDDC data centers to provide comprehensive information on NOAA scientific videos and related products. The paper also explains the library's standard procedures on digital video conversion, preservation and archiving.

I. INTRODUCCION

It has always been challenging for special libraries to provide public access to scientific data and information that exists in both physical and digital formats, via geospatially enabled digital objects that are a result of scientific explorations and research. In order to provide easy discovery and access to this scientific information, librarians need to challenge themselves with metadata creation, standardization, and exchange. NOAA Central

Library (NCL) staff, NOAA scientists, and IT specialists have embarked on a pioneering collaboration in metadata exchange to accomplish these tasks, under the Library's Video Data Management System (VDMS) project.

II. NOAA LIBRARY COLLECTIONS: AN OVERVIEW

Located in Silver Spring, Maryland, the NCL networks with over thirty NOAA libraries nationwide. NCL is considered to hold the historically richest and most comprehensive multidisciplinary scientific collection in hydrographic surveying, oceanography, ocean engineering, atmospheric sciences (climatology and meteorology), meteorological satellite applications, living marine resources, geophysics, cartography, and mathematics in the United States. The NOAA Library Network collections are unique: over 40% of the items in NOAALINC (the online catalog) and in WorldCat are not found anywhere else. The NCL also has original media collections from a variety of NOAA scientific investigations, explorations, and research.

In addition, information on the unique and historically valuable NOAA ocean exploration documents, scientific data and products, in the forms of digital videos, still images, and datasets, have also been entered into NOAALINC, the National Oceanographic Data Center (NODC) Ocean Archive System (OAS), and other oceanographic information catalogs and databases. These additional resources are available thanks to the Library's collaboration with other NOAA projects and programs, such as the Video Data Management System (VDMS), Climate Data Modernization Program (CDMP), and NODC Cruise Report Project.

III. VDMS HISTORY

The NCL staff has been collaborating with NOAA's Office of Ocean Exploration and Research (OER), NODC, and National Coastal Data Development Center (NCDDC) to address the requirements for archiving, preserving, managing, and providing online access to digital videos and still images from OER oceanographic expeditions since late 2002.

The VDMS developed to facilitate managing online information and access to video and images obtained during OER Signature oceanographic expeditions. This system now enables access to hundreds of digital video clips, highlight movies, still

images, and related documents and products from OER expeditions. Using discovery tools available via NOAALINC catalog (<http://www.lib.noaa.gov/>), NOAA scientists and other researchers can discover and download online video and still images for such phenomena as deep sea coral or polar ecosystems areas. Upon special arrangement, NCL can provide equipment and an appropriate environment for users to view, copy and/or download requested off-line scientific video data, or to view original expedition tapes preserved in the NOAA Library Archives. Additional online information includes related cruise reports, educational lesson plans, original video and image annotation documents, digital maps and Web sites, and links to other oceanographic observation data.

IV. METADATA CHALLENGES

During the development of the VDMS project plan, the VDMS team defined and established several 'best practices' to support OER video data management requirements. The VDMS team collaborated with NODC and NCDDC teams to develop a set of tools, documents, and workflows to insure standardized access to data and information collected during NOAA OER-sponsored cruises. These include:

- A VDMS requirements document that defines technical standards for the system, describes archival storage formats and conditions, and specifies online retrieval requirements;
- Metadata standards requirements based upon the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata and the international library standard, MARC21 (Machine Readable Catalog for 21st Century). Other systems, such as CoRIS (Coral Reef Information System), may also access VDMS information in NOAALINC these same metadata protocols;
- A crosswalk and converter in NCDDC's MERMAid (Metadata Enterprise Resource Management Aid) system [1] that enables sharing common metadata in both FGDC and MARC21 records by converting FGDC metadata to MARCXML format;
- Collection-level (parent) MARC21 metadata records from the NCL that can be linked to activity-level (child) FGDC metadata records created by the NCDDC using the crosswalk and conversion techniques

developed for the VDMS (see Fig. 1). These metadata records provide the descriptive information for both NOAAALINC and MERMAid.

- Collections, tapes, clips, highlights videos, or still images may be documented using guidelines developed by the VDMS working group. These guidelines are referred as DV12 (Digital Video 12 descriptive elements) for video, and DI12 (Digital Image 12 descriptive elements) for still images. The DV12 and DI12 templates were developed to help field personnel document the contents of videos they created. Information from the field personnel (in DV12 or DI12 form) are used by NCL staff to create MARC21 collection level records.
- A Video Lab created by the VDMS team provides a venue for transferring video data from original media to mass-storage media. It is equipped with video processing hardware and software that enables VDMS project staff to develop and document workflow processes, enables more robust file conversion and video management capabilities, and creates files suitable for online access and long term retention. Final Cut Pro 5™ software is installed on two Apple™ G5 video editing workstations to allow in-house video processing by encoding raw video data into online-accessible formats. [2]

To help scientists and data managers in the field create complete descriptive metadata about their data collections, and to assure the minimum information for metadata creation in both scientific FGDC and library MARC standards, the Library developed simple metadata templates - guidelines for digital video (DV12) and digital still images (DI12). In DV12 and DI12 templates, the data managers are asked, in twelve core elements, to provide the basic information about the captured video and images. These include the official name of expedition, cruise or project, date, name(s) of the PI(s), abstract, subject coverage, geographic coordinates, altitude or depth, type of investigations, type of data collected, format and size of the video footage, video highlights and still images, system requirements for direct and remote access, video and image annotations, and contact information.

(http://docs.lib.noaa.gov/OEDV/VDMS_DOCS/DV_12_05112004.doc),

(http://docs.lib.noaa.gov/OEDV/VDMS_DOCS/DI_12_7052006.doc)

Scientists, librarians and archivists then use this information to create MARC21, FGDC, or Dublin Core metadata records. The DV12 and DI12 guidelines request scientists to provide the core information for the technical and descriptive metadata elements.

This additional metadata is needed to ensure that observations can be referenced to a specific point in the world's ocean. Geographic information is often collected automatically from shipboard systems, using a Global Positioning System (GPS) or other navigation technology. The geospatial relationship of video data to positional information is maintained by using time-stamp information available from both video and navigation sources. In addition to time-stamp annotations for individual video tapes, data managers and Principal Investigators often provide descriptive metadata for a video and/or image collection using the aforementioned DV12 and/or DI12 templates. They may also provide copies of cruise reports and/or other data reports created concurrent with or subsequent to the video collection. Descriptive information about video content, technical details about file formats, encoding algorithms, and processing equipment are needed to ensure that these videos are accessible and meaningful for the long term.

The backbone for the VDMS data discovery system is NOAAALINC, the library's online catalog. The VDMS Team organizes the videos, still images and related products received from OER oceanographic expeditions into physical and online collections to create the collection-level metadata record in MARC 21 format. Through the descriptive metadata, the high level information on scientific exploration and media products are preserved. Technical metadata reflecting the media format and size, codec type, frame size, encoding speed and similar information are provided.

In MARC21 metadata, the set of URLs provides online access to different types of digital documents, including video clips and highlights, still images, PDF, Microsoft Word and Excel documents. The library metadata also provides direct links to related scientific databases and data access points, including the NODC OAS

(<http://www.nodc.noaa.gov/Archive/Search/>),

(http://www.ncddc.noaa.gov/website/google_maps/OE/mapsOE.htm),

or CoRIS (<http://coris.noaa.gov/>) for further discovery of the raw observations, cruise investigations in these GIS driven systems, or simply for additional research information.

V. HOW CHILDREN MEET THEIR PARENTS

It was a turning point for the VDMS project when the Library of Congress developed the MARCXML schema to work with MARC21 data in the XML environment. The Library offers cruise information via its MARC21 collection level metadata, and the NOAA scientific data centers, including NGDC, NODC and NCDDC, generate and offer the same information via their FGDC metadata. In other words, the Library creates the parent-type metadata record for each OER expedition, after which the scientific data centers create a series of child-type metadata records for each cruise-related document, data set and scientific data product. From these sequential activities, an idea emerged to complement, rather than duplicate, each other's efforts in metadata creation. Several months of discussion and the Library's customization of the FGDC- to-MARC21 (see Fig. 1) mapping resulted in implementing the metadata converter in MERMAid, NCDDC's metadata catalog. A Cruise Information Management System (CIMS) application, developed for OER by NCDDC, provides a user interface to capture the metadata for data-producing activities, events and operations on OER-sponsored expeditions. In post-cruise processing, CIMS generates FGDC metadata records within MERMAid. MERMAid then exports the metadata records bound for the NCL into the MARCXML standard format. The converter supporting the crosswalk between FGDC and MARCXML outputs the series of MARCXML metadata records describing one video tape or related products from the particular OER expedition.

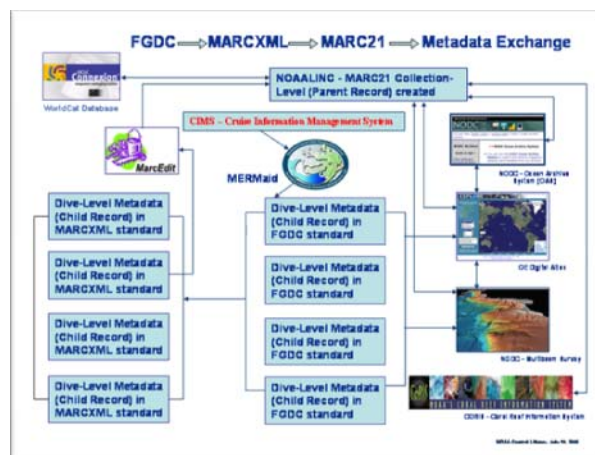


Figure 1. FGDC—MARCXML—MARC21 metadata conversion and exchange workflow

http://docs.lib.noaa.gov/OEDV/VDMS_DOCS/MARCXML2008.pdf

Next, the MARCXML files generated from MERMAid are sent to NCL for further and final conversion using MarcEdit [3], a free MARC conversion and editing software utility, developed by Terry Reese of Oregon State University (<http://oregonstate.edu/~reese/marcedit/html>). This very useful program helps the Library team to turn the MARCXML files into MARC21 file format, and even do some global editing before the records are entered into the NOAALINC database to meet their parent or, the collection-level, metadata record. Each child, or product-level, record includes the linking entry with the parent, aka collection-level, record. Currently, the NCDDC team is considering the use in MERMAid of the Library's collection level records, converted via MarcEdit, to FGDC standards.

Since the converted records from FGDC to MARCXML require manual editing once entered into NOAALINC, the NCDDC programmers are currently working on improving the conversion process. The NCL metadata records are created according to a set of standards, guidelines, and controlled vocabulary, including MARC, AACR2 (Anglo-American Cataloging Rules 2nd ed.), LCSH (Library Congress Subject Headings), LCNA (Library of Congress Name Authority), and LCC (Library of Congress Classification). The scientific FGDC metadata is created according to its own set of rules and authorities, including GNIS (Geographic Names Index System), PLSS (Public Land Survey System), FIPS (Federal Information Processing System location codes), ISO 19115 (Theme Keyword, and Topic Category), to name a few.

The NCDDC programmers are working to enhance the metadata input in CIMS, to make it library-standards friendly and to assure the quality conversion process from FGDC to MARCXML. Upon completion of the enhancement, less manual or macro-driven editing in NOAAALINC, the library online catalog, will be required.

VI. MULTIPLE ACCESS TO OCEANOGRAPHIC METADATA

Many types of data and information products are collected and created during a typical OER oceanographic cruise, including planning documents, cruise summary reports, laboratory specimen lists, video and still images, and navigation and other observational data. Most of the products and media are received in the NCL by the VDMS team for further processing.

The primary objectives of VDMS are: to provide timely online access to OER video and images; to educate the general public about NOAA oceanographic expeditions and underwater investigations; to archive and preserve unique video and related data for the long term; to foster collaboration between NOAA librarians, scientists, and data managers from different NOAA line and program offices; to use or extend existing library tools, guidelines, and metadata standards to support new media formats, such as digital videos, still images, and digital text documents; and to enhance data access and metadata sharing between the NCL's NOAAALINC, NODC's OAS and NCDDC's MERMAid catalogs, CoRIS and OER Digital Atlas applications.

To meet these objectives, the quality and informative collection- and product-level metadata is essential. To support the library users in the process of discovering oceanographic data, the Library provides comprehensive or so called "fully loaded" metadata information on oceanographic expeditions, including online access to a variety of cruise products and scientific data sets (Figure 2).

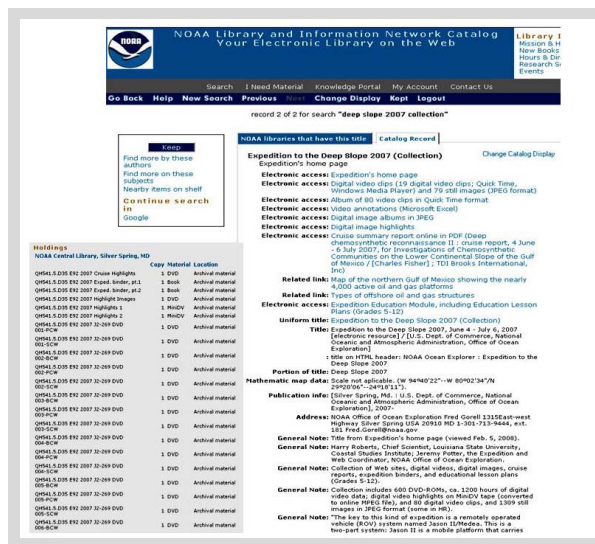


Figure 2. Example of NOAAALINC metadata record in MARC21 standard format. Collection-level record describes Deep Slope 2007 expedition.

The metadata in NOAAALINC offers access to oceanographic online products that reside on the library server as well as to the NODC's OAS and OER Digital Atlas databases.

The media and documents from each expedition are archived, and maintained in the NCL. The raw oceanographic observations are stored, preserved and archived in NODC. NODC provides long term archival storage, management and stewardship of digital oceanographic observation data and metadata. As illustrated in Figure 3, most ocean data archived at NODC can be discovered and downloaded using the NODC OAS online at

<http://www.nodc.noaa.gov/Archive/Search/>



Figure 3. NODC OAS's Accession record provides access to raw observations from the polar cruise described in NOAAALINC

However, the expedition information, including scientific observations, dive activities, and ship navigational data can be tracked down and retrieved via the OER Digital Atlas developed by the NCDDC team using Google™ Maps applications. Users can filter the color-coded dots by year and exploration theme on the map. Each dot represents an OER signature cruise from 2001 to the present. Figure 4 illustrates the OER Digital Atlas interface.



Figure 4. OE Digital Atlas using Google Maps provides the NOAA signature expeditions information
http://www.ncddc.noaa.gov/website/google_maps/OE/mapsOE.htm

This project provides an ongoing opportunity to improve the quality and completeness of metadata and information used in the NOAAALINC catalog and the NODC OAS and to provide online access to NOAA ocean exploration video and related data to a global customer base.

Currently, there are 38 digital collection-level (parent) metadata records and over 800 product-level (child) records in NOAAALINC. These records provide off-line access to over 3000 original video tapes and discs. The links also provide online access to more than 500 clips and highlights, as well as over 8000 still images and hundreds of related documents. These translate into 50.7 GB of video and related products posted on the Library server (<http://docs.lib.noaa.gov/OEDV/>), organized into 729 folders, and containing 19, 629 video and related files. According to our monthly statistics, there were over 261, 324 online visits to these folders and over 2.31 TB of data were downloaded, since October 1, 2008. That confirms that there is a great demand within NOAA and the general public for access to scientific oceanographic information in video and image formats.

The success of the VDMS project demonstrates that much has been done, but there is more work to do. As additional resources become available, plans include providing online access to broader subsets of available digital video holdings, hosting an informal seminar series, and examining how other groups (e.g., educators, other scientists) use the digital video data. The VDMS project working group continues to collaborate closely with NOAA OER project scientists, oceanographers, and IT specialists to develop data management requirements and strategies. [2]

The VDMS Project's future plans include increasing access to the multi-platform oceanographic information systems through NOAAALINC and the WorldCat catalog. WorldCat is the world's largest and richest database of bibliographic information, linking over 200 million metadata records from the catalogs of over 54,000 libraries in 109 countries. In collaboration with NCDDC, the teams plan to develop a web-based portal from which diverse OER ocean data, including video, still image, and audio files will be accessible via text-driven searches or from map-driven searches using a digital atlas.

VI. SUMMARY

Currently, the VDMS team is preparing the software, equipment and workflow capabilities to discover, access and preserve the forthcoming high-definition video and image data from the new NOAA Ship *Okeanos Explorer*.

VII. ABBREVIATIONS

ATDB – NODC Accession Tracking Data Base
CoRIS – NOAA’s Coral Reef Information System
CIMS – Cruise Information Management System
FGDC – Federal Geographic Data Committee
GPS – Global Positioning System
MARC21 -- MACHine Readable Cataloging
(Standards for the representation and communication of bibliographic and related information in machine-readable form)
MARCXML – MARC21 eXtensible Markup Language (XML schema to represent MARC21 data elements.)
MERMAid -- Metadata Enterprise Resource Management Aid
NCL – NOAA Central Library
NODC – National Oceanographic Data Center
NOAALINC – NOAA Library and Information Network Catalog
NCDDC – National Coastal Data Development Center
OAS – NODC Ocean Archive System
OER – NOAA Office of Ocean Exploration and Research
OCLC – Online Computer Library Center, Inc.
PI – Principal Investigator
VDMS – Video Data Management System

REFERENCES

- [1] US Department of Commerce, National Oceanic and Atmospheric Administration, National Oceanographic Data Center, National Coastal Data Development Center, Metadata Enterprise Resource Management Aid (MERMAid), <http://www.ncddc.noaa.gov/metadatarresource/metadadata-tools> (Last accessed on July 23, 2009).
- [2] Fiolek, A., D. W. Collins, 2008. Video Data Management System Archives, Preserves and Provides Online Access to NOAA Deep-Sea Corals Video and Image Data. In: *Oceans 2008 MTS/IEEE, Quebec City: oceans, poles and climate: technological challenges*. Columbia, Md.: Marine Technology Society; Piscataway, N.J.: IEEE. CD-ROM. Also available online from the NOAA Central Library at: http://docs.lib.noaa.gov/OEDV/VDMS_DOCS/Oceans2008_paper_final.pdf (Last accessed on July 23, 2009).
- [3] Reese, Terry, MarcEdit version 5.1, free software available online at <http://oregonstate.edu/~reese/marcedit/html/>. (Last accessed on July 23, 2009).