

NOAA Ocean Exploration Digital Video and Image Data: Archiving, Preserving, and Accessing Online Oceanographic Information

Anna Fiolek, Dorothy Anderson, Janice Beattie; NOAA Central Library; Silver Spring, MD; Donald W. Collins; NOAA National Oceanographic Data Center; Silver Spring, MD

Abstract

To address the increasing requirements for archiving, preserving and managing digital video, still images, and audio resources, the National Oceanic and Atmospheric Administration's (NOAA) Office of Ocean Exploration (OE) embarked on the Video Data Management System (VDMS) Pilot Project, in collaboration with the National Oceanographic Data Center (NODC), National Coastal Data Development Center (NCDDC), and the NOAA Central Library (NCL). Since 2002, the OE Integrated Product Team (IPT) has been developing a standardized capability for archiving these disparate types of data and information.

NCL staff led the development of the Video Data Management System (VDMS) Project Plan, which is a part of a larger comprehensive OE Data Management Project. The VDMS team was asked to define and establish 'best practices' to support OE video data management requirements. They developed metadata guidelines for digital video (DV12) and digital still images (DI12) to help scientists and data managers in the field create complete metadata about their data. These guidelines also facilitate creation of MARC21, FGDC, or Dublin Core standard metadata records. They also proposed a work-flow for managing 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.

The VDMS Pilot Project presently manages offline access to more than 1500 MiniDV and 500 DVCAM tapes, over 1000 DVDs, and online access to more than 100 digital video clips and highlights collected during NOAA ocean exploration cruises. Currently, access to the NOAA cruise video highlights and related documents is provided through NOAALINC, the NCL online catalog at <http://www.lib.noaa.gov>. A growing collection of digital data obtained during OE cruises, including video, still images, and in situ ocean observations, is archived at NODC. These data are accessible through the search and retrieval functions of the NODC Ocean Archive System (OAS) at <http://www.nodc.noaa.gov/Archive/Search/>.

The OE VDMS Pilot Project has demonstrated its initial capability to acquire, document, manage, preserve and provide access to digital video and still image data. Five-year VDMS Project plans (2006-2010) include:

- Increasing access to multi-platform video images through the NOAA Libraries Online Catalog (NOAALINC) and the Online Computerized Library Center (OCLC) WorldCat catalog.
- Developing a web-based portal from which diverse OE 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.

- Using the NODC Archive Management System as the central digital file management repository for video, still images, ocean observations, and related documentation.

- Expanding the scope of relevant video and image data to include similar data and information from other NOAA Line Offices and Program Offices.

Introduction

Many programs and offices of the National Oceanic and Atmospheric Administration (NOAA) routinely create digital and analog videos that document program activities and data collection tasks (e.g., submersible dives, short highlights clips). As required by NOAA Administrative Orders NAO 15-217 and NAO 205-17, the NOAA Central Library (NCL) and National Oceanographic Data Center (NODC) are receiving an increasing number of video data collections from diverse NOAA components, including the Office of Ocean Exploration (OE), the National Marine Sanctuaries Program (NMSP), and the Coral Reef Conservation Program.

During a typical oceanographic cruise, many types of information and data are developed, including planning documents, cruise summary reports, laboratory specimen lists, video and still images, and navigation and other observational data. Beginning in late 2002, the NCL and NODC began collaborating with OE data managers to develop and implement an end-to-end data management plan for data and information collected during ocean exploration cruises. The Integrated Product Team (IPT) was formed to develop a comprehensive plan, with several working groups to focus on components of the overall plan. One working group was tasked with developing a Video Data Management System (VDMS) for acquiring, accessioning, cataloging, maintaining and providing access to digital video data. The IPT recognized that it would also be beneficial if the requirements, documentation and system could serve as a model for the whole agency.

Video and still image data present many challenges for Principal Investigators (PIs), data managers, and metadata librarians who work with the video and images after the conclusion of the data collection project. This paper describes many of the processes implemented by NOAA to assure that digital video data files from these and other sources are managed consistently and effectively for the long term with minimal staff resource requirements.

Media Management

The primary media currently used for capturing video images are MiniDV, DVCAM, and VHS tape media or directly onto DVD. Each of these media types uses a different native encoding (file format structure) to create moving images. Video processing software is required to convert (encode) native video formats (i.e. uncompressed .dv, or .avi) to current industry and archival standard formats (e.g., MPEG-2 or MPEG-4) to facilitate online access and for long-term management. At present, NCL and NODC use MPEG-2 as the archival standard encoded format. Online access formats include MPEG-2, QuickTime™, Windows Media Player™, and RealMedia™ formats.

At present, NCL manages a growing collection of multiple video media. This collection of original media includes more than 1500 MiniDV tapes, 500 DVCAM tapes, 245 VHS tapes and more than 1000 DVDs. These original media contain the entire sequence of original video footage obtained and provide a relatively complete record of events during a cruise, submersible dive or other activity. Original video media are currently stored in a climate controlled room and will be migrated to new media as necessary for ongoing long-term archival preservation.

In addition to original media, NCL archives clips and highlights created from the original, full length raw video. Clips typically contain very short (15-60 seconds) excerpts of interesting or unusual features. Highlights are usually a series of short video segments (2-15 minutes) selected by the PI and/or data manager as a representative sample of images collected during the cruise. NCL provides online access to clips and highlights video through links in the MARC21 records of NOAAALINC, the library online catalog (<http://www.lib.noaa.gov/uhtbin/Webcat>). A search for “digital video online” will list all catalog metadata records containing one or more links to digital videos, as well as other related media and documents. Figure 1 shows an example of NCL metadata record in MARC21 standard. As additional resources become available, online access to the contents of original tapes may be implemented.

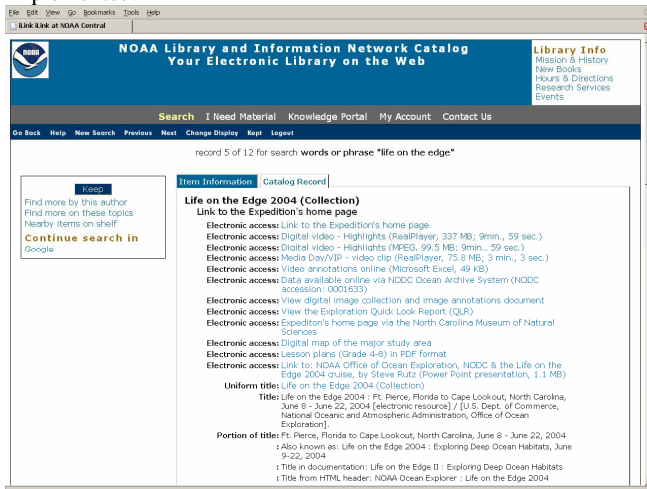


Figure 1. Example of NOAAALINC metadata record in MARC21 standard format. Collection-level record describes Life on the Edge 2004 Expedition.

An accurate inventory of still images, usually copied from CD-ROM or DVD, is more difficult to obtain: for example, a single collection of still images captured during one OE cruise is comprised of more than 90,000 images. Some still image collections are being prepared for inclusion in the NCL Photo Library, an online image collection arranged into subject matter ‘albums’ (<http://www.photolib.noaa.gov/>).

Video Data Management Tools

The VDMS working group developed a set of tools and products to assure harmonized and standard access to valuable data and information collected during NOAA OE-sponsored cruises.

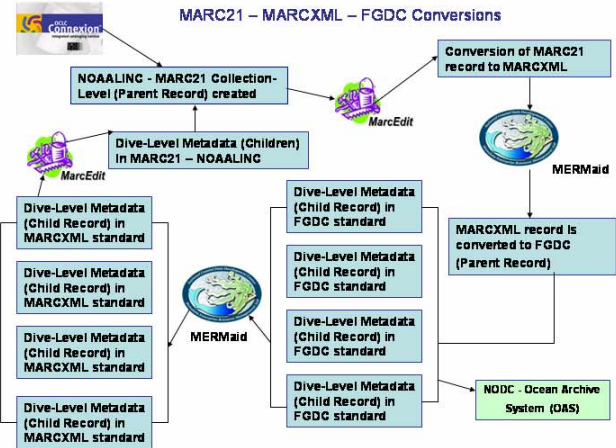


Figure 2. MARC21 to MARCXML to FGDC conversion work-flow.

NVDS products and tools include:

- VDMS requirements document defines technical standards for the system, describes archival storage formats and conditions, and specifies online retrieval requirements.
- Metadata standards requirements include using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata to describe geospatial scientific data and MARC21, the library-wide standard for documenting a resource (e.g., video or still image). Databases supporting FGDC and MARC21 metadata records may be accessed by the VDMS simultaneously using the Z39.50 search protocol.
- A crosswalk was developed and implemented to enable sharing common metadata in both FGDC and MARC21 metadata records. A MARC21 collection-level (parent) metadata record created and converted to MARCXML, using MarcEdit [1] is sent to the National Coastal Data Development Center (NCDDC), where they are converted to FGDC records, using MERMAid [2] for the OE online catalog. Conversely, FGDC dive-level metadata (child) records containing additional metadata elements related to individual dive tapes are sent to the NCL for inclusion in NOAAALINC (Figure 2).
- Each collection, tape, clip, or highlights video may be documented using guidelines developed by the VDMS working group. These guidelines are referred to as DV12

(Digital Video 12 descriptive elements) for video and DI12 (Digital Image 12 descriptive elements) for still images.

- 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.
- Video processing hardware and software will be available soon for improving workflow processes and enabling more robust file conversion and management capabilities.

Documentation Requirements

Imagine trying to understand a video created today in 50 years: When and where was the video taken, by whom, and for what reason? What is the format of the file and the encoding used to create the file? These few questions highlight the need for obtaining as much descriptive and technical metadata from the PI or data manager soon after the completion of each cruise. Video shot during a submersible dive is often the primary data collection activity for the dive and is intrinsically a form of geospatial data. As a result, the use of video as a source for quantifiable geospatial data (e.g., percentage of specific seafloor area covered by sponges and echinoderms, identification of common and unique species at a specific location) makes the content management and metadata requirements somewhat different than video that is primarily a record of a historic event.

Additional metadata is needed to assure that observations can be referenced to a specific point in the world ocean. Geographic information is often collected automatically from shipboard systems, using Global Positioning System (GPS) or other navigation technologies. To maintain the geospatial relevance of video data, video footage is typically matched to navigation information by using time-stamp information available from both the video and navigation sources. In addition to time-stamp oriented annotations for individual video tapes, data managers and PIs often provide descriptive metadata for a video and/or image collection using the DV12 and/or DI12 templates. They also typically provide copies of cruise reports and/or other data reports that were developed concurrent with or subsequent to the video collection. Descriptive information about the content of a video and the technical details about file formats, encoding algorithms, and processing equipment are needed to ensure that these videos are accessible and meaningful to future generations. Navigation data, technical details and reports are archived with other observational data at NODC.

VDMS Archival Processes

When video tapes and related data are received at the NCL, the metadata librarian creates a MARC21 record in the NOAA LINC online database for the collection of video and related documents. Figure 3 illustrates the generic flow of video and other data from data collectors and originators to the archive centers [3]. NCL notifies NODC that a new collection of video (and other materials) has been acquired. An NODC data content manager creates an accession entry in the NODC Accession Tracking Data Base (ATDB) [4] for the collection of tapes and related materials.

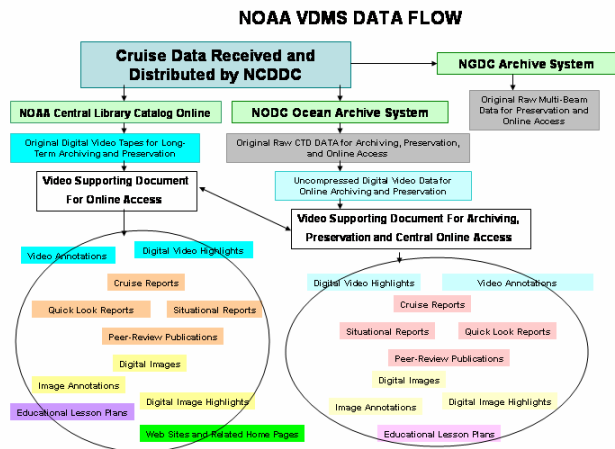


Figure 3. Schematic diagram showing generic data flow in the NOAA VDMS (full size image available online at http://docs.lib.noaa.gov/OEDV/VDMS_Workflow_Ocean2005.jpg).

NODC provides long term archival storage, management and stewardship of digital oceanographic data and metadata. Each new collection of data is assigned an NODC accession number as a tracking number for the collection. As shown in Figure 4, each digital accession has the same basic structure, which is intended to identify and separate files from the originating source (e.g., video, cruise reports, oceanographic observations) and files created by NODC about the original files [4]. A copy of clips and highlights files from specific OE cruises is placed in the associated NODC accession, with a link to the file established in the NCL MARC21 record for the collection (Figure 1). Non-video observation data collected during the cruise with video data may also be placed in the same NODC accession or a separate accession, with a reference to the video accession.

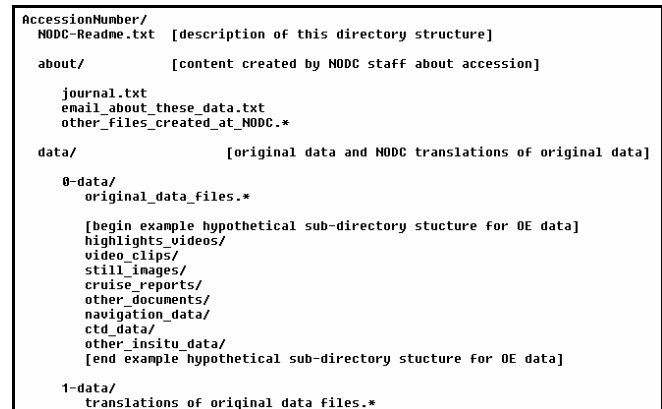


Figure 4. Generic structure of each digital accession in the NODC Archive Management System.

Archival digital files maintained at NODC are stored primarily on RAID media, with offsite backups on tape media [4]. At present, original video media provided by cruise PIs or data managers are stored and maintained by the NCL. Other hard-copy information, including a binder of paper forms created during a cruise, are also stored and maintained by NCL. As resources for

digitizing these paper media become available, digital surrogates of paper forms will be maintained in the NODC digital archives with related ocean observation data.

Most ocean data archived at NODC can be discovered and downloaded using the NODC Ocean Archive System (OAS, online at <http://www.nodc.noaa.gov/search/prod/>) [4]. FGDC metadata are automatically harvested from the NODC ATDB accession entry for inclusion in the NODC Metadata Manager and Repository (NMMR) database.

Summary and Future Activities

The VDMS project provides a good foundation for establishing NOAA procedures to assure that the agency's scientific video data in both physical and online formats is archived and preserved for future generations. The VDMS project working group continues to collaborate closely with NOAA OE project scientists, oceanographers, and IT specialists to develop data management requirements and strategies. This project provides an ongoing opportunity to improve the quality and completeness of metadata and information used in the NOAA LINC catalog and NODC Ocean Archive System and to provide online access to NOAA ocean exploration video and related data to a global customer base.

The successes of the VDMS Pilot Project demonstrate that much has been done, but there is more work to do. The NOAA Central Library and NODC will soon bring two video processing workstations on line, thus improving in-house video processing capabilities to enable encoding raw video data into online-accessible versions. As additional resources become available, other plans include providing online access to broader subsets of available digital video holdings, hosting an informal seminar series that highlights video collections, and examining how other groups (e.g., educators, other scientists) are using digital video data from NOAA. Long term VDMS Project plans include:

- Increasing access to multi-platform video images through the NOAA Libraries Online Catalog (NOAALINC) and the Online Computerized Library Center (OCLC) WorldCat catalog. WorldCat is the world's largest and richest database of bibliographic information, linking approximately 60 million bibliographic records from the catalogs of over 53,500 libraries in 96 countries.
- Developing a web-based portal from which diverse OE 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.
- Using the NODC Archive Management System as the central digital file management repository for video, still images, ocean observations, and related documentation.
- Expanding the scope of relevant video and image data to include similar data and information from other NOAA Line Offices and Program Offices.

An online 'tour' of the VDMS Project is available at http://docs.lib.noaa.gov/OEDV/VDMS_DEMO_2005/noaavdmsdemo.rmvb. The RealMedia™ demo file is approximately 140 Mbytes in size and takes about 20 minutes to play. RealPlayer™ is required to play this file.

Abbreviations

MARC21 (MACHine Readable Catalog) – Standards for the representation and communication of bibliographic and related information in machine-readable form.

OCLC – Online Computer Library Center, Inc.

MARXML - Simple XML schema which contains MARC21 data elements.

References

- [1] Reese, Terry, MarcEdit version 4.5, Free software available online at <http://oregonstate.edu/~reese/marcedit/html/>. (Last accessed on 6 March 2006).
- [2] US Department of Commerce, National Oceanic and Atmospheric Administration, National Oceanographic Data Center, National Coastal Data Development Center, Metadata Enterprise Resource Management Aid (MERMAid). Free software available online at <http://www.ncddc.noaa.gov/Metadata>. (Last accessed on 6 March 2006).
- [3] Fiolek, A., D. W. Collins, D. Anderson, J. Beattie, NOAA Video Management System (VDMS): no video, no Oscar@! Proc. Oceans 2005, Abstract #050410-01, in press (2005).
- [4] D.W. Collins and S.B. Rutz, The NODC archive management system: archiving marine data for ocean exploration and beyond. Proc. Oceans 2005, Abstract #050520-02, in press. (2005).

Author Biographies

Anna Fiolek serves as a Metadata Librarian at the NOAA Central Library since May 1994. She holds her MA in Linguistics from the University of Wroclaw, Poland, and MLS from the University of Maryland. As Metadata Librarian, she creates master metadata records to all NOAA and NOAA-related documents in all formats, including online resources. She trains and supports other cataloging librarians in the NOAA Library network and serves as a metadata specialist for several NOAA-wide projects.

Dorothy Anderson received a BA in French from The Ohio State University and her MLS from The Catholic University of America and. She is currently the Head Reference Librarian at the NOAA Central Library where she has worked for 12 years. She has been the Public Services lead of the Video Data Management System working group since its inception. Ms. Anderson serves on various matrix teams within NOAA and is an active participant in developing and maintaining virtual libraries.

Janice Beattie is Director of the NOAA Central and Regional Libraries which houses one of the largest marine and atmospheric collections in the country. The NOAA Library network serves NOAA and the public nationwide. Ms. Beattie is the U.S. representative to the U.N. Aquatic Sciences and Fisheries Board, which is responsible for the publication of the ASFA database. She is also a member of the Group of Experts on Marine Information Management (GEMIM), a working group within the Intergovernmental Oceanographic Commission (IOC).

Donald W. Collins received a BS in Geology from Indiana University, an MS in Oceanography from the University of New Hampshire, and an MLS from the University of Maryland. He has been an oceanographer at the National Oceanographic Data Center since 1988. He started out at NODC providing reference services to a diverse constituency of research scientists, business people, engineers and the public. In 1998, he began working on improving data cataloging and metadata, revising controlled vocabularies, and updating archival practices for NODC data collections.