

The undeniable key to success in discovering video is metadata and a metadata search portal into which they are published. NCEI utilizes a combination of Geoportals based on the granularity of the metadata. A “collection” level Geoportal houses metadata for collections of data – in the case of OER video, the cruises on which video was recorded. A “granular” level Geoportal houses metadata for individual data files – in the case of OER video, the video segments, the video streams, and the video highlights. NCEI’s collection-level geoportal is accessed at the following URL: <https://data.nodc.noaa.gov/geoportal/catalog/search/search.page>.

Once a Geoportal user has discovered a collection, and followed the link to the Details in the expanded results, the user will be on the “landing page” of that collection. In the case of OER video, there is a link on the Access tab that will take the user to the OER Video Portal (the result of this VDDMMI Project) for that collection. Additionally, the user can follow the links to the low-resolution video data via either an https or an ftp protocol. This would be useful if the user is interested in all of the videos available for that cruise. If, however, the user is looking for something specific in the video, the OER Video Portal provides a way to search the video collection using specified criteria.

The cruise metadata record, and the subsequent NCEI landing page for the cruise accession, provides information about the cruise and links to data and information resources. The granular video metadata records, and the subsequent NCEI landing pages for the video resources, provide the name and abstract for each video segment and links to a previewable version of the video and a link to place an order for the corresponding full-resolution version. But in order to get to these landing pages, the user needs to be able to discover them – and that is done through the metadata.

A key technique for OER video data management is to embed critical metadata in the filename of each asset. This insures that files will always be understandable and discoverable even if the link to metadata files are lost. This requires a commitment to a standardized naming convention, and makes for some long file names. Using the naming convention OER created for *Okeanos Explorer* video, it is possible to identify from the file name:

- the cruise during which the video was taken, expressed as expedition id (two digit ship code + two digit field season year + two digit sequential cruise number within field season, and optionally an ‘L’ and a two digit sequential leg number)
- the dive number (for the Okeanos Explorer, this is a unique number within a cruise. The Okeanos uses DIVE01, DIVE02, etc for each cruise)
- the time stamp associated with the first frame of the video, expressed as year, month, day, time in hours, minutes, and seconds, relative to UTC
- the camera and platform from which the video was taken
- codes and free field text describing the content of the video
- the resolution of the video – “low” for low-resolution. If this attribute is missing, the file is the source or full-resolution.
- the file format container - .mov is used for the full-resolution and .mp4 is used for the low-resolution.

For example, the video file captured during cruise EX1402, Leg 3, on April 12, 2014 at 19:10:15 (UTC) using the High Definition camera on the main ROV submersible, showing a mound and a shrimp would be named

EX1402L3_VID_20140412T191051Z_ROVHD_MOUND_SHI_Low.mp4. The “Low.mp4” suffix indicates that the video file is a low-resolution video file with an .mp4 wrapper. The corresponding full resolution video segment would be named identically, but without the “_Low” – EX1402L3_VID_20140412T191051Z_ROVHD_MOUND_SHI.mov.

Another stage of metadata generation requires some information about the cruise on which the ROV operations are occurring. The OER Data Management Team uses an Access database called a Cruise Information Management System (CIMS) to capture much of the metadata needed. Metadata is gathered about the cruise at the planning stage and saved in CIMS. The CIMS then outputs an ISO collection-level metadata record with the “who, what, when, where, why, and how” of the cruise. Once the cruise is completed, the data metrics are entered for each data type collected, including the video. CIMS will export certain header type information in an xml format which is then used by video annotation extraction routines and metadata generation routines. The following list of sources of metadata for each video segment include:

- Cruise header information (cruise title, abstract, dates, principal investigator, ocean basin, ocean sub-basin, geospatial boundaries, mission themes)
- Dive observation notes (dive purpose, dive observation, geospatial boundaries, chief scientist, dive site name, maximum depth)
- Embedded video metadata (duration, wrapper, codec, bitrate, frame rate)
- Video file name (UTC date/time stamp, vehicle, camera, taxa codes, free form text)
- Submersible CTD (establish range values for depth, temperature, and salinity using start time of video segment and duration)
- Submersible Navigation (latitude, longitude, altitude, attitude, pitch, roll, heave)
- Scientific Chat Log (video annotation by participating scientists using code tables and free form text)
- If the video segment captures a specimen collection, the EX Sampling Operations Database Application (SODA) (genus/species, size, condition, weight, collector, comments)
- If the video segment goes viral on social media, the names (in all conceived spellings, of the viral monikers)
- If the video segment contains a frame grab used in the Okeanos Benthic Animal Guide, the Group, Subgroup, Category, Subcategory, Family, Taxonomic ID, Phylum, Subphylum, Superclass, Class, Subclass, Infraclass, Superorder, Order, SubOrder, Infraorder, Superfamily, Subfamily, and Subgenus of the imaged biological individual.

A resource the team relies heavily on is NCEI’s Docucomp utility found at <http://www.ngdc.noaa.gov/docucomp>. Docucomp is an ISO metadata component library that allows account holders (the reader can request an account through Docucomp) to build metadata

snippets that tend to be relatively static. Universally Unique Identifiers (UUIDs) can be generated through online UUID generators and then included in the metadata snippet before saving. Such snippets include CI_ResponsibleParty metadata segments for organizations and for personnel; MI_Instrument metadata segments for instruments; MI_Platform for vessels and submersibles; and MD_Keywords for standard vocabularies. These are then referenced in the metadata records as calls to the component library. These “unresolved” records are then resolved through the Docucomp before being saved and published. This will take a snapshot in time of the details for each of these metadata snippets. If information changes for any of them, the updates are made in Docucomp and future metadata records will get the new information. If older metadata requires the new information, re-resolving will perform that function.

All of these data are then mapped into the ISO 19115-2 formatted metadata template designed and refined over the course of the project:

Video Segment Metadata Template:

https://www.ncddc.noaa.gov/oer-waf/media/templates/EX_Video_Segment_Template.xml

Video Segment Metadata Example:

https://www.ncddc.noaa.gov/oer-waf/media/templates/EX1605L3_VID_20160625T025344Z_ROVHD_ROC_RIDGE_FSH.mov.xml