VIDEO DATA MANAGEMENT BEST PRACTICES

For any reader with video data assets looking for a similar solution, OER encourages adoption of any methods described and materials offered through this effort. The team also recommends several best practices that will help ensure a successful transition to a capable, extensible, technology-leveraged video data management system. In order to aid technology transfer, the team hopes that some of these lessons learned may help:

- *Open access to scientific data is a target state that should drive development*
    - The OER video data collection represents a high volume, high value data set, comprising a wide product range and a diverse and wide user base with different requirements. To maximize the value of this dataset, all users should have open access to this data.
    - The 2015 NOAA plan to increase Public Access to Research Results (PARR), written in response to the White House Office of Science and Technology Policy (OSTP) Memorandum of February 2013, requires public accessibility to all data and publications produced by federal researchers or by recipients of federal funds, such as the OER video data collection.
    - Open access is leveraged by advances in technology, evolution of metadata standards, and social uses of scientific data; therefore, it will be necessary to continually revisit development of the tools created to achieve it.
- *The widespread availability of rich metadata is essential for a successful, unified method of managing a large video data collection*
    - Reliable, complete geospatial, temporal, environmental and keyword-based metadata are necessary for self-service discovery and access
    - Metadata must be gathered starting early in the planning stages and continuing through the entire lifecycle of video asset use
    - Metadata should be associated with both collection-level and granule-level structures in a hierarchical model that allows users to traverse easily between the two
- *Video should be captured at the highest quality levels with the lowest level of compression consistent with available resources and expected downstream processing and use*
    - *Okeanos Explorer* video is captured at 1920 x 1080i employing the ProRes 422 video codec inside a QuickTime container format, at a bit rate of 145 Mbps. This format utilizes mild compression but is still considered broadcast quality.
    - SMPTE Timecode referenced to UTC should be embedded in the video signal, fed from an aboard-ship master clock to synchronize all video sources during dives.

    - SMPTE Timestamp should be made part of the video filename to aid discovery and to enable metadata extraction from other sources
- *Video storage and preservation requires a large capacity infrastructure that can handle current and future needs*
    - Move video files to stable storage media as soon as possible to avoid deterioration
    - Long term storage and preservation of *Okeanos Explorer* video is housed in the NOAA CLASS infrastructure, which has enough capacity to meet current and expected demand for the scientific and broadcast communities.
    - CLASS data is backed up frequently to insure long term preservation of archived video.
    - To support discoverability, accessibility, and usability, online streaming or downloading of preview copies is served from spinning disk. These copies are also stored in CLASS for long-term preservation and backup.
    - A well-designed middleware geoportal should allow for immediate preview of online video and short-term ordering of near-line video.
- *Standards-based approaches ensure interoperability*
    - Metadata conforms to ISO 19115-2 standard for image-based geospatial data. This facilitates the hierarchical linking of each granule-level video segment through its metadata record to the parent collection which is the cruise during which the video was captured.
    - Each cruise is assigned a stable, persistent Digital Object Identifier (DOI), minted by the California Digital Library's EZID system. Assigning a DOI to OER video ensures that it can always be discoverable, even if the URL linked to it changes for any reason.
- *Pilot projects on a small scale before committing to full-scale implementation*
    - Information learned during the FADGI study gave the VDMMI team the confidence that video capture was robust and that it was time to move onto the storage and dissemination phase.
    - The Amazon Cloud pilot gave the VDMMI team the confidence that the core concept of two tiers of storage, near-line tape and online spinning disk, linked by a middleware geoportal, was sound. However, it was critical that the storage system have 100% preservation durability in order to fulfill NOAA's mission of stewardship for the future.
    - External presentations of the system in the last few months before rollout were critical to test out the viability of the new system before other oceanographers and data scientists.
    - Feedback gained at the June 2016 presentation to the Underwater Video Workshop held at the University of Rhode Island was especially helpful, as the

audience was largely comprised of peers holding positions at other underwater video archives, such as the Scripps Institution of Oceanography, Woods Hole Oceanographic Institution, Ocean Networks Canada, NASA Ames Research Center, National Science Foundation, and the Schmidt Ocean Institute.