

EXPERIENCES DEVELOPING OAIS-RM RECOMMENDED SUBMISSION AGREEMENTS

Robert H. Rank¹, Constantino Cremidis², Scott McCormick³, Jeremy Throwe⁴

NOAA/NESDIS/OSD/CLASS¹, CSC², Global Science & Technology (GST)³, Science and Technology Corporation (STC)⁴

ABSTRACT

The National Environmental Satellite, Data, and Information Service (NESDIS) under the National Oceanic and Atmospheric Administration (NOAA) is responsible for the collection, archival, and dissemination of environmental data collected by a variety of in situ and remote sensing observing systems operated by NOAA and by a number of its partners, e.g., National Aeronautics and Space Administration (NASA). To prepare for large increases in its data holdings, the NESDIS Office of Systems Development (OSD), has been developing the Comprehensive Large Array-data Stewardship System (CLASS). CLASS currently provides data acquisition, storage, access, and dissemination at three distinct locations: the NOAA Satellite Operations Facility (NSOF) in Suitland, Maryland; the National Climatic Data Center (NCDC) in Asheville, North Carolina; and the National Geophysical Data Center (NGDC) in Boulder, Colorado.

CLASS currently holds data from Polar-orbiting Operational Environmental Satellite (POES), Geo-stationary Operational Environmental Satellite (GOES), Defense Meteorological Satellite Program (DMSP), and Initial Joint Polar-Orbiting Operational Satellite System (IJOSS) satellites and derived products. Planned future data campaigns will add data from the Next Generation Weather Radar (NEXRAD) system, as well as the National Polar-orbiting Operational Environmental Satellite System (NPOESS), NPOESS Preparatory Project (NPP), Earth Observing System Moderate-resolution Imaging Spectroradiometer (EOS/MODIS), and GOES-R series satellite systems.

CLASS has adopted the recommendations of the Open Archival Information System Reference Model (OAIS-RM) and has developed processes around these recommendations. The recommendation with the most impact on the way that CLASS conducts business is the development of Submission Agreements between the Producers and the Archive.

CLASS is in a position not defined within the OAIS-RM. NOAA has designated the three NOAA National Data Centers (NNDs), namely NCDC, National Oceanographic Data Center (NODC), and NGDC, collectively as the official NOAA archive. CLASS is responsible for

implementing the IT components of the Archives. Because of this special arrangement, Submission Agreements cannot be developed solely between the Producer and the Archive (as represented by the data center designated as responsible for the particular data), but must include CLASS, an organization independent of but related to the Archive. This three-party Submission Agreement has implications especially because, from the Producers' point of view, the Producer is negotiating a Submission Agreement with one entity – the Archive. However from CLASS' and the Data Center's respective points of view, there are three parties involved in the development and approval of the Submission Agreement. This paper will present the approach, processes, and experiences that CLASS has had while developing Submission Agreements under this three-party arrangement.

Another special situation, not necessarily envisioned by the developers of the OAIS-RM, is that many Producers can share a single IT infrastructure for submitting data. For example with CLASS, all Producers associated with the Environmental Satellite Processing Center (ESPC) use ESPC's Data Distribution System (DDS) as the mechanism for sending data to CLASS. Against this reality, it is impractical for all Submission Agreements to include (and replicate) the technical details about how the DDS will make data available to CLASS. In addition, it is common in the Producer's organization for the group responsible for generating the data to be a very different group from the one that makes the data available to CLASS. CLASS therefore decided to develop Interface Control Documents (ICDs) to accompany one or more applicable Submission Agreements. ICDs are developed and agreed to between CLASS and the group inside the Producer's organization responsible for delivering the data to CLASS. This paper will make the case, based on CLASS' experience developing Submission Agreements and ICDs, that some Archives might find it necessary to develop both Submission Agreements and ICDs.

Finally, this paper will present CLASS' experiences with Producers while developing Submission Agreements, and the benefits to the overall program (on the Producer's side), that were drawn from the development of these Submission Agreements. CLASS' experiences have shown that Submission Agreements are not only a way to

Index Terms— OAIS-RM, Submission Agreements, ICD, Producer, Archive CLASS

Over the past few years, NOAA has decided to move towards an enterprise IT solution in support of NOAA's Archives. As part of this evolution, NOAA has identified CLASS as the system that provides IT capabilities to support NOAA's archive mission. CLASS will partner with the NOAA National Data Centers (NNDCs) to help them support NOAA's archive mission. [1]

Due to the diversity of the NOAA mission; NOAA information preservers exhibit significant variation with regard to mission, domains, management processes, and designated communities. It is because of such diversity that it is not expected that there will be a single 'NOAA Archive'. Instead, there will be multiple 'NOAA Archives', each performing its own archive administration and preservation planning activities while sharing the IT services of enterprise systems like CLASS. These various 'Archives' will share some preservation planning and management policies and procedures.

To better support the development of a NOAA Archive vision and to ensure comparable functions and procedures across NOAA Archives; NOAA has started an OAIS-RM functionality mapping effort. This effort is being led by the NNDCs and CLASS and has as its goal a consensus identification of which functions and responsibilities fall under the IT component of the Archive and which fall under the Archive itself. The completion of this functionality and responsibility mapping will contribute toward development of joint preservation planning and administration policies and procedures; ensuring a basic level of uniformity across NOAA's Archives.

A Submission Agreement is an agreement between a producer and an archive on how the data will be submitted to the archive for preservation and dissemination to the designated community. Submission Agreements include information on: human contacts (technical, metadata); designated community data access and dissemination options; data transfers; protocols; validation; errors and actions; data formats and standards; metadata; data quality information; lineage; parameters; and system performance.

2. MULTIPLE NOAA ARCHIVES

Submission Agreement in the context of the NOAA Archives is a three-party effort. This adds an additional complexity not envisioned by the developers of the OAIS-RM. It is under this reality, that NOAA is developing various new policies, processes, and procedures to address these complexities. In particular, CLASS and the NNDCs have developed and implemented a number of standard processes to aid in the development of Submission Agreements. The five most important are:

1) The development of a Submission Agreement template. This is a comprehensive and fully annotated template that provides clear examples for all sections of the Submission Agreement document. This template is frequently reviewed and updated to ensure usability, clarity, and understanding.

2) Establishment of a single Point of Contact (POC) for the producer with the Archive. One issue that causes frustration from the producer's perspective is having to deal with multiple people during the development of a Submission Agreement - and thus having to repeat concepts every time the submission agreement reaches a milestone causing a POC to change. The NNDCs have established a 'cradle to grave' approach when developing a new Submission Agreement. A Submission Agreement is assigned to one person who is fully responsible for coordinating with the producer and CLASS for the completion and approval of the Submission Agreement. Following NNDCs' lead, CLASS has adopted a similar procedure by assigning a single person responsible for supporting the development of that Submission Agreement.

3) Establishment of monthly Submission Agreement status review meetings among CLASS and the representatives of NOAA archives who are working on Submission Agreements. These coordination meetings ensure that CLASS and the Archives are aware of which Submission Agreements are under development, what the pertinent need dates are, and what the producers' and designated community's expectations are, among other issues.

4) Establishment of a Metadata Management Repository and a Metadata Manager. A single system and a single organization, the NOAA Metadata Management Repository (NMMR), is responsible for gathering 'collection level metadata' [4] for NOAA archives and for ensuring uniformity of format and content. This single organization interprets the metadata standards and provides guidance on which information, in what format, and to which level of detail must be collected in the Submission Agreement.

5) Improvements to the requirements management process. It was noted that during the development of Submission Agreements, especially when drafting the sections for data access and dissemination, additional functional requirements were sometimes being included. Support for these access and dissemination requests would mean that the IT system supporting NOAA archives would need to be enhanced, in some cases at significant cost and/or with little forewarning. CLASS recognized that the development of a Submission Agreement in some instances had the unintended consequence of circumventing the requirements vetting and approval processes already in place. Enhancements to the Submission Agreement review and approval process now include: clear and early identification of proposed new requirements; compilation of these requirements into an appropriate document; and submission of that document through the standard requirements review and approval process. These steps are necessary before proceeding with review and approval of the Submission Agreement that brought the need for these requirements to light.

By implementing these five process improvements for the development of Submission Agreements, CLASS in particular and NOAA's archives in general have found that many of the early adoption difficulties have been overcome and that the process of developing Submission Agreements is smoother and better accepted by producers.

4. NEED FOR INTERFACE CONTROL DOCUMENT

Groups of NOAA's data producers, especially its satellite data processing systems, tend to share data distribution points. A diverse group will typically share a single data production and distribution system. This is true for the current Environmental Satellite Processing System (ESPC) where different producers share the same IT systems; and will be true in the future for the NPOESS Data Exploitation System (NDE) and for the NPOESS Interface Data Processing Segment (IDPS), among others.

Under these circumstances, it is impractical to include in every Submission Agreement the technical characteristics of the shared interface through which the data will be submitted to CLASS for archival storage. The reasons are: (a) one change to the interface would require changes to multiple documents; (b) producers do not have authority to negotiate the interface requirements for this IT system; and (c) NNDCs do not have authority to negotiate CLASS' interface requirements.

To address this problem, CLASS implemented the development of an ICD for documenting the technical characteristics of each interface. CLASS and the representatives of the IT system that will submit the data to

CLASS develop and control this ICD. The ICD describes: data transfer protocols; data transfer validation; error conditions; recovery mechanisms; data transfer volumes and performance requirements; network architecture; and system security considerations.

The ICD is developed by and for the IT systems. It is written without representation of members from the NNDCs (as part of the NOAA Archives) and the producers. Although, in theory, one Submission Agreement can be related to multiple ICDs, our experience, so far, has been that the opposite is true - one ICD relates to multiple submission agreements. These two documents, a Submission Agreement and its corresponding ICD, are related by the data that they document.

5. BENEFITS TO PRODUCERS

During the development of Submission Agreements for new data flows that will be submitted to CLASS for storage, archival, and dissemination, CLASS' experience is that producers tend to be, understandably, more interested in resolving issues related to algorithm development, data production, and dissemination to real-time users than in working with NOAA Archives to develop a Submission Agreement.

It is because of this that NOAA Archives have found it necessary to clearly state the goals and benefits to producers for development of Submission Agreements. Without full participation from the producer, CLASS has found that development of a Submission Agreement takes a prolonged amount of time (years) to complete. The following are the primary benefits that a producer will receive by developing a Submission Agreement:

Early identification of required metadata. During the development of Submission Agreements, all collection and granule level metadata that will be utilized by the archive to catalog, provide access to, and disseminate the data to the designated community must be identified. Required metadata is clearly identified, helping the producer minimize last minute changes or scope creep, by asking the following questions: (a) How will the data be cataloged? (b) How will the data be discovered and searched? (c) How will the data be disseminated to the designated community? (d) What additional information is valuable to the designated community? and (e) What additional information is required for reprocessing of the data?

Early identification of data that needs to be submitted to the archive. This benefit is primarily a consequence of the previous benefit. The full needs of which data should be archived are quickly identified by asking questions like: (a) What is the primary use of the

data by the designated community? (b) What information is required for reprocessing the data? and (c) What companion data is needed for understanding the data? It is CLASS' experience that it is very easy to fall into the trap of saying that 'all' data will be submitted for archival without clearly defining what 'all' means. Development of a Submission Agreement is the surest and most efficient way of clearly and completely defining 'all data'. The development of the Submission Agreement has the additional benefit of defining the interface performance requirements and therefore supports system capacity planning activities.

CLASS has identified many other benefits to the producer. The two presented here are simply the ones with the most significant beneficial impact for the producer. Experience while developing the Submission Agreement for support of the Jason-2 mission clearly demonstrated that the producer was greatly benefited by early development of the Submission Agreement.

6. SUMMARY

In this paper we presented how a complex and diverse organization such as NOAA greatly benefits from the adoption of standard processes and procedures in general and from the adoption of the OAIS-RM in particular. CLASS, as the enterprise IT system in support of NOAA archives, has embraced the OAIS-RM and its recommendations. One such important recommendation is the development of Submission Agreements between the producer and the Archive.

We also presented the unique challenges that NOAA and its Archives face when developing Submission Agreements. These challenges include the fact that more than two organizations develop and approve the Submission Agreements, and that producers may share a single IT system for submitting their data to the Archive. We presented some of the procedures that the NNDCs and CLASS have implemented to address these challenges.

We concluded by presenting, based on CLASS' experience, the most important benefits that a producer will attain through development of the Submission Agreement.

7. REFERENCES

- [1] R. Rank, "Enterprise IT Support for NOAA Archives", IGARSS 2007
- [2] D. Zirkle - 'CLASS Target System Architecture, Topics', unpublished
- [3] R. Rank, K. McDonald - Use of a standard process for submission of data for archive and preservation - IGARSS 2006
- [4] R. Rank, A. Kidd - 'CLASS plans for providing archive, access and distribution services for GOES-R data' - SPIE 2006