JP8.4 EVOLUTION OF NOAA'S SATELLITE ARCHIVE: TRANSITION OF CLASS DEVELOPMENT TO OPERATIONS AT NOAA'S DATA CENTERS AND FUTURE PLANS

John Bates*
National Climate Data Center, Asheville, North Carolina

Ken Casey
National Oceanographic Data Center, Silver Spring, Maryland

Eric Kihn
National Geophysical Data Center, Boulder, Colorado

ABSTRACT

NOAA's Data Centers, the National Climatic Data Center, the National Oceanographic Data Center, and the National Geophysical Data Center, have the unique responsibility for the long-term management and stewardship of the bulk of NOAA's data. These data are widely used in addressing today's and tomorrow's environmental issues. In order to cope with the major increase in data volume from Earth remote sensing in the near future, NOAA initiated an information technology infusion project, the Comprehensive Large-Array Stewardship System (CLASS), to provide the needed hardware and software for data ingest and storage. This development system is now being transitioned to operations within NOAA's Data Centers who are ultimately responsible for archiving and servicing the data.

1. BACKGROUND

In the 1990's, both NOAA and NASA created digital satellite data archives to ingest, store, and provide access to customers. Due to the high volume of data, and the complexity of the hardware and software required, these archives were centrally developed. As a result, these digital archives evolved on a unique path separated from the institutions in which they were housed. This led to a proliferation of digital archives for earth science information since these centrally managed archives coexisted within institutions that already had existing archives unique to their particular scientific specialties. At the time, these digital satellite archives proved to be a boon to researchers who had heretofor only limited access to large volumes of satellite data for research and applications.

Over time, as the satellite data records became longer and more mature, the applications of satellite information broadened to include investigators well beyond experts in remote sensing. This, in turn, has created a need to accommodate customer search and access tools that can be more easily used across many different archives. This led to efforts in both NOAA and NASA to move to a concept of Co-operative and Federated archives and the transition of the digital satellite archives back into the local institutions. This has been accompanied by a re-examination of the roles of centralized software development and hardware acquisition, where savings can occur when common needs are addressed, versus decentralized archives that can be more responsive to customer needs in particular thematic areas.

2. THE OPEN ARCHIVE INFORMATION SYSTEMS REFERENCE MODEL (OAIS-RM)

The rapid expansion and proliferation of digital archive across a wide variety of disciplines led to the recognition that there was no common framework, in terms of terminology or functional areas, to share the knowledge of archival that was being gained. This led to a movement, initially by space and earth science archive communities, to adopt into identified the functional areas common to digital archives. In order to define this so-called reference model, what is meant by archiving of data had to be defined and then this digital archiving process had to be broken down into a few common functional areas.

The resulting reference model (CCSDS, 2002) is identified as the open archive information system. This reference model is targeted to several categories of potential users including archive designers, archive users, archive managers, and standards developers. An important aspect of the reference model is that it has adopted common terminology that crosses various disciplines such as traditional archivists, scientific data centers, and digital libraries. This reference model has since become widely adopted as a starting point in digital preservation efforts.

The reference model provides a framework for understanding and applying concepts needed for long-term digital information preservation. In this case long-term means long enough to be concerned about changing information technology on which the digital information is stored. It provides a set of minimal responsibilities to distinguish and OAIS from other uses of the term archive. These responsibilities include that the OAIS:

- Negotiates and accepts information from information producers

* *Corresponding author address:* John J. Bates, NOAA National Climatic Data Center, 151 Patton Ave., Asheville, NC 28801; e-mail: John.J.Bates@noaa.gov

- Obtains sufficient control to ensure long-term preservation
- Determines which communities (designated) need to be able to understand the preserved information
- Ensures the information to be preserved is independently understandable to the Designated Communities
- Follows documented policies and procedures that ensure the information is preserved against all reasonable contingencies
- Makes the preserved information available to the Designated Communities in forms understandable to those communities

3. OAIS-RM FUNCTIONAL AREAS AND IDENTIFYING ROLES AND RESPONSIBILITIES IN COOPERATING EARTH SCIENCE ARCHIVES

The definition of information within an OAIS provides the most fundamental basis for examining the roles and responsibilities of common information technology activities versus the role of subject matter experts. Digital information is always expressed by some form of data and thus there is a critical role or information technology expertise in the movement and storage of these data. Just as critical, however, is the need for these data to be interpreted by its representation information that yields information objects to be preserved. The due to the complexity of earth science information, the identification and preservation of representation information by subject matter experts is also a key component in digital archives.

The OAIS identification of functional entities within a digital archive is being used by NOAA's data centers in their effort to move the CLASS system from development to operations. The functional areas illustrated in Figure 1 include:

- **Ingest:** This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers and prepare the contents for storage and management within the archive
- **Archival Storage:** This entity provides the services and functions for the storage, maintenance and retrieval of Archival Information Packages
- **Data Management:** This entity provides the services and functions for populating, maintaining, and accessing both descriptive information that identifies and documents archive holdings and internal archive administrative data.
- **Administration:** This entity manages the overall operation of the archive system
- **Preservation Planning:** This entity monitors the environment of the OAIS and provides recommendations to ensure that the information stored in the OAIS remain accessible to the Designated Community over the long term.
- **Access:** This entity supports Consumers in determining the existence, description, location and availability of information stored in the OAIS and allows Consumers to request and receive information products.

The initial mapping of the CLASS IT system from development to operations within NOAA's data centers involves mapping the full OAIS functional areas to support the concept of cooperating archives as illustrated in figure 2. This involves standardizing the submission (or ingest – Ing) and dissemination (or access- Acc) methods to the core CLASS IT system. This will allow the subject matter experts at the data centers to provide their expertise where it is most effective; in the details of the needs of the consumers who are accessing the data. Similarly, as CLASS evolves to accommodate a wider variety of data sets, the subject matter experts at the Data Centers are also the experts on what data and metadata are needed from the producers to ensure full information preservation.

4. WHAT SHOULD NOAA ARCHIVE

Recently, the National Research Council (NRC) completed a study (NRC, 2007) requested by NOAA to provide high-level advice on how to archive and provide access to the broad range of environmental data NOAA and its partners collect. This report identified nine principles for NOAA to follow in terms of its archive and access policies. Of particular interest to the archive and the CLASS IT project is principle 7, 'A formal, ongoing process, with broad community input, is needed to decide what data to archive and what data not to archive'. The report went on to state that, 'NOAA needs to establish a high-level, enterprise-wide approach to decide what data to include in their archives. The decision to archive (or not to archive) data should be driven by societal benefits, should ensure that irreplaceable environmental data are preserved, and should explicitly incorporate broad community engagement and coordination with other agencies.'

In response, NOAA has developed a procedure for scientific records appraisal and archival approval that are in the process of being formally adopted. As the NRC report also noted, 'although it is impossible to save everything, the goal of NOAA's data management enterprise should be to ensure that the broadest possible collection of environmental data is archived and made discoverable and accessible to the widest possible range of users'. A key part of ensuring this is the case is the development of an appraisal procedure as outlined below (Figure 3). Details of the process will be available once the policy is formally adopted by NOAA. Broadly speaking, there are 4 steps proposed in this appraisal: 1) identify the records, 2) appraise the records, 3) a decision/approval process, and 4) implementation of the decision. There are proposed provisions for the approval process to happen at different levels depending on the size, cost, and complexity of the data sets being considered.

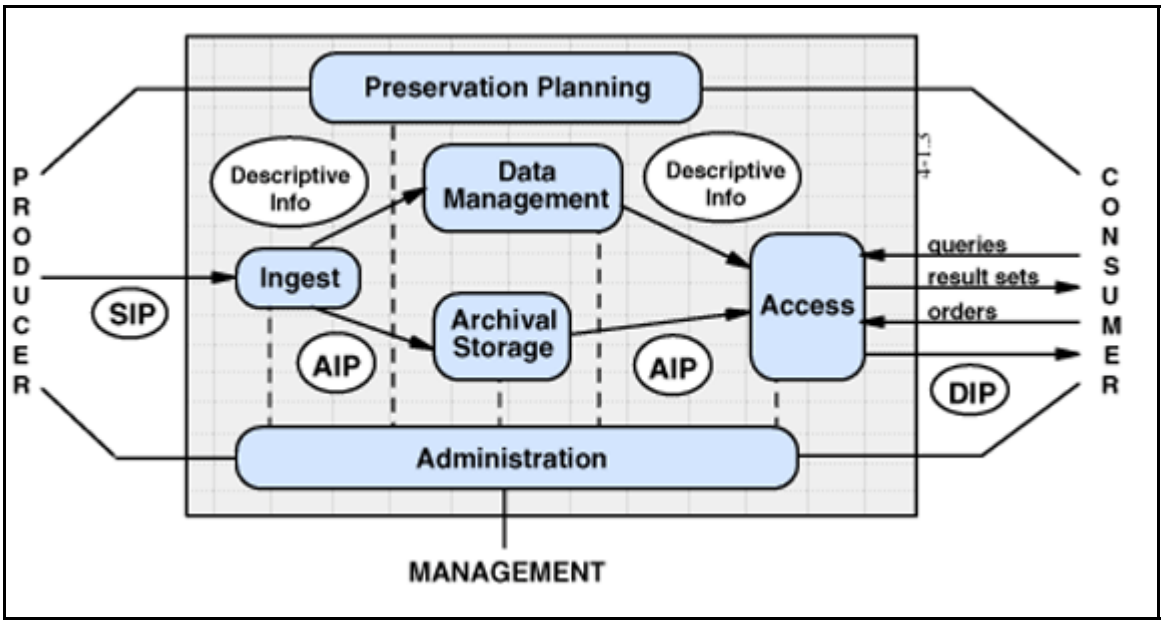
5. CONCLUSIONS

NOAA's data centers and archive IT infrastructure are evolving to meeting the changing needs of customers and to take advantage of technology improvements. An important component of this evolution is the transfer of the CLASS IT ingest, archive, and access system from development to operations within NOAA's data centers. This will allow for improved efficiency and customer service as the data centers can take better advantage of the common services the CLASS system provides and provide improved user experience by being more responsive to customers.

6. REFERENCES

1. CCSDS (Consultative Committee for Space Data Systems), 2002: *Reference Model for an Open Archival Information System (OAIS).* CCSDS 650.0-B-1. Blue Book. Issue 1. CCSDS Secretariat, Washington, D.C., 148 pp. http://public.ccsds.org/publications/archive/650x0b1.pdf

2. NRC, 2007: *Environmental Data Management at NOAA – Archiving, Stewardship, and Access*. National Academy Press, Washington, D.C., 116 pp.



SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

Figure 1. Functional areas of an open archive (from reference 1).

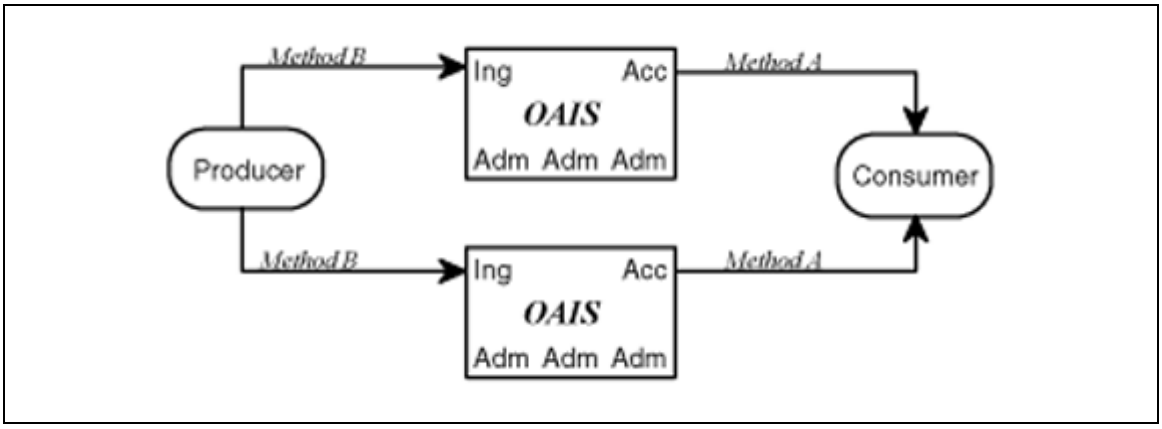


Figure 2. Notional concept of cooperating archives (from reference 1).

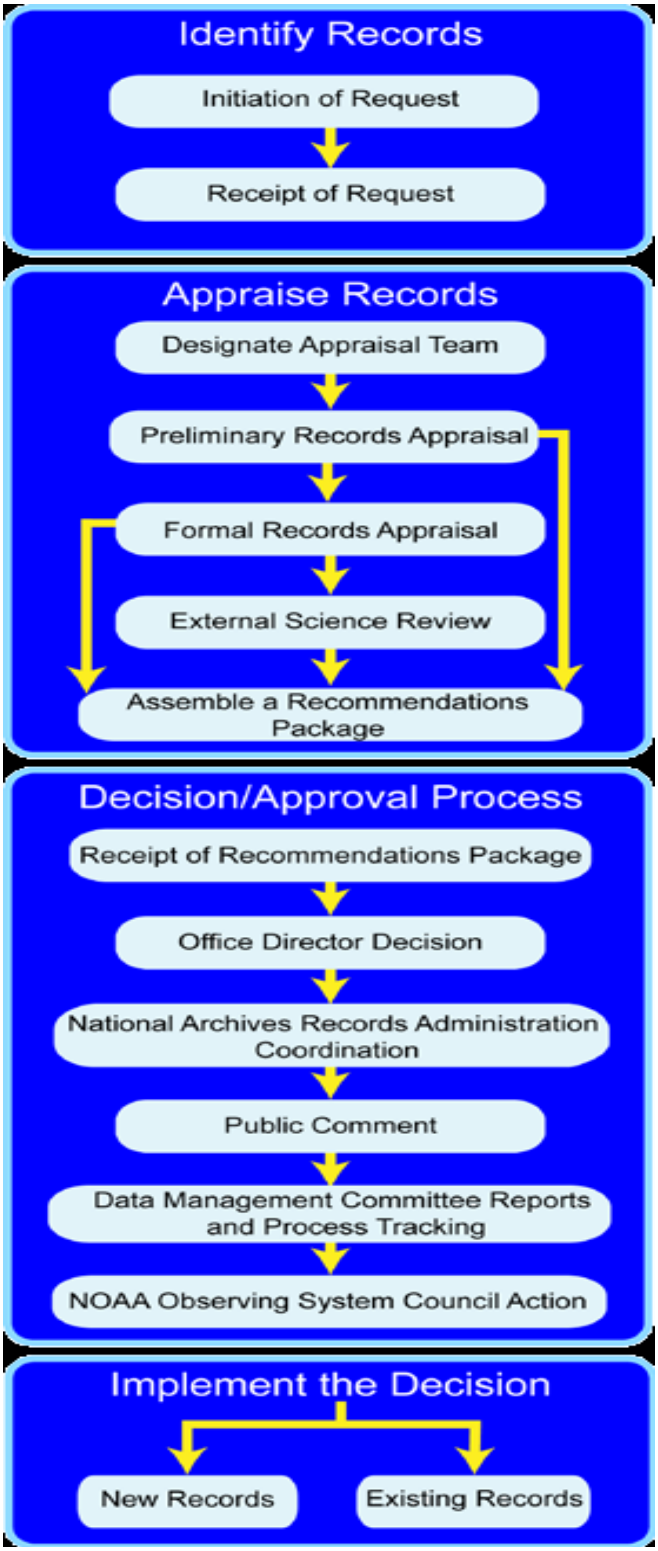


Figure 3. Notional schematic of the proposed procedure for scientific records appraisal and archival approval.