CGMS

CGMS-34, NOAA-WP-19
Prepared by NOAA
Agenda Item: II/2
Discussed in WG2

REQUIREMENTS FOR ARCHIVED SATELLITE DATA TO BE USABLE FOR CLIMATE STUDIES

In response to CGMS Action 33.05

NOAA-WP-19 offers NOAA's perspective on the steps needed to make archived satellite data usable for climate studies. The scientific aspects of this issue were summarized in NOAA's input to CGMS XXXII Action 30.09 and so this response will focus on specific archival issues that address long-term preservation and interoperability of satellite data. NOAA recommends CGMS members explore adoption of the Open Archival Information System Reference Model (OAIS-RM) as a framework for long-term satellite information preservation.

Requirements for Archived Satellite Data to be Usable for Climate Studies

1 Introduction

This purpose of this paper is to provide NOAA's perspective on the steps needed to make archived satellite data usable for climate studies. The scientific aspects of this issue were summarized in NOAA's input to CGMS XXXII Action 30.09 and so this response will focus on specific archival issues that address long-term preservation and interoperability of satellite data. NOAA recommends CGMS members explore adoption of the Open Archival Information System Reference Model (OAIS-RM) as a framework for long-term satellite information preservation.

2. OAIS Reference Model

The OAIS-RM is an ISO standard that describes the functions of an archive and suggests a generic framework for packaging data and metadata so that the long-term survivability of their information content is enhanced. The OAIS-RM is widely recognized and generally accepted in a variety of digital libraries, including the space agency community. Use of this standard will promote interoperability by providing a standard vocabulary for archive services, including the Submission Agreements that constitute a "quasi-contractual" view of the agreement between an archive and a data provider ("producer," per the OAIS-RM.) By using Submission Agreements, an archive can improve its cost effectiveness, its planning, and provide better service to both its producers and its user communities ("consumers"). The OAIS-RM also includes provisions for monitoring "designated communities" of consumers, thereby encouraging archives to better tie their accession plans and services to real user community needs.

NOAA is implementing the OAIS-RM as a collaboration between NOAA's enterprise long-term storage and access system, the Comprehensive Large-Array data Stewardship System (CLASS), and its National Data Centers. The OAIS-RM does not specify a particular implementation, and NOAA recognizes that different members will have different requirements. NOAA's implementation strategy is that CLASS will build the IT infrastructure supporting ingest, data management, archival storage, access, and administration, while the data centers will assume the archive responsibilities described in the OAIS-RM, including preservation planning. The data centers will use CLASS's IT infrastructure in support of fulfilling their archive responsibilities.

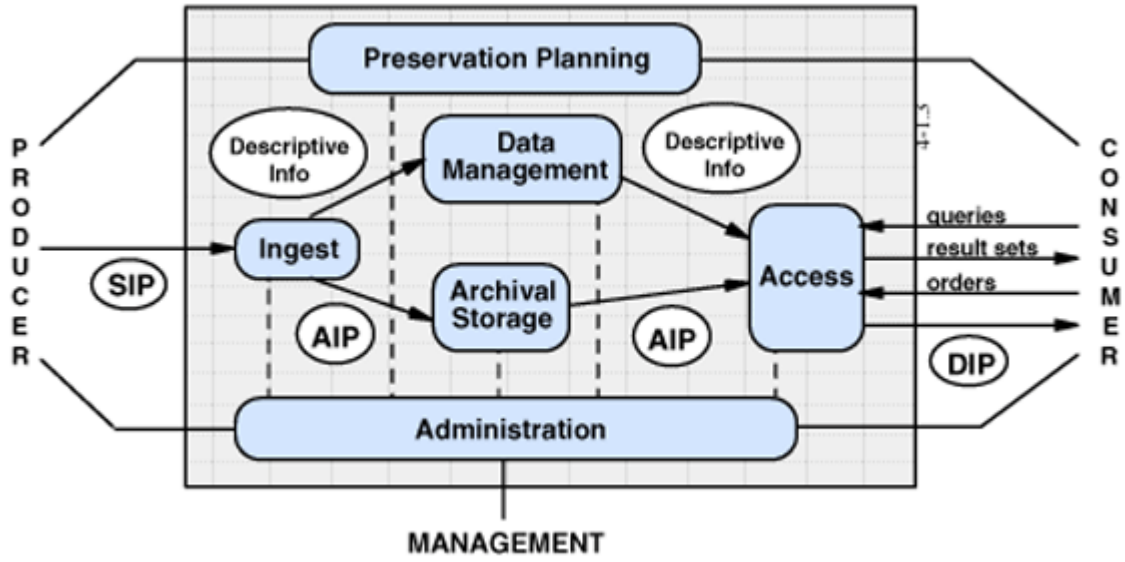

Figure 1. Diagram from the OAIS-RM showing the functional entities and information flows that characterize an archive.

In addition, the data centers and CLASS will work to provide tools that facilitate ease of use and interoperability. The keys to providing interoperability, both among NOAA's data holdings and among CGMS members, are service oriented architecture and use of common data models. These components are summarized in the next sections.

3. Service Oriented Architecture

In order to provide international and interagency interoperability, the NOAA recommends CGMS members move toward a Service Oriented Architecture (SOA), through which advancements in machine-to-machine interfaces can act to reduce the overall cost of operations. In the U.S., such a SOA approach is recognized in the Federal Enterprise Architecture (FEA), which is intended to increase the interoperability of U.S. Information Technology (IT) systems. There are similar approaches in other portions of the international IT systems, such as the Grid computing initiatives being undertaken by the E.U. as part of their next five year planning framework. The Integrated Data Environment for Geospatial Data (GEO-IDE) is an effort to move beyond a discipline-specific view of atmospheric, oceanic, land surface, and space observations to a coordinated, efficient, integrated, and interoperable environment. This effort is intended to approach data management from both a top-down and a bottom-up design, using standards and structured data types as key elements, while using Web services as an operational "glue". While the GEO-IDE effort intends to create a "system of systems" that will allow users to seamlessly discover, manipulate, visualize, and fuse data for a wide variety of purposes, the developers also recognize that the effort cannot be efficient and cost effective if it attempts to build a new (and very expensive) infrastructure that disrupts current services and data provision to users. As a result, there is a requirement to

work with legacy systems and to view the new services as an evolutionary convergence of heritage systems and new capability.

The Common Data Model (CDM), an effort being led in the U.S. by the National Science Foundation's University Corporation for Atmospheric Research (UCAR), holds great promise for enhancing interoperability of data sets commonly used by CGMS members within an SOA. A data model describes data objects and the operations used to access and manipulate them, in a manner similar to object oriented programming. The Common Data Model will provide a common interface to the NetCDF and HDF file formats and also allow for inclusion of legacy formats, such as historical satellite level 1b. In this way, the Common Data Model will abstract underlying format details and enable the user to focus on using—not accessing—the information. Additional layers of the common data model define standards for coordinate systems and scientific data types. NOAA recommends that CGMS members explore the potential of the Common Data Model for enhancing interoperability of current, future, and legacy data sets.