The case for document preservation

Many NASA Earth Observing System (EOS) missions have either reached the end of their active life or are nearing it. In order to ensure that future users can draw maximum benefit from the data for years to come, it is necessary for us not only to preserve the data properly, but also the preservation of all associated metadata, calibrations/valid data, and important documentation or other artifacts. Missing any information that could hamper the Data-Information-Knowledge chain takes away from the ability of future generations of researchers to draw value from our data. Preservation of data products is a fairly well-defined task for the NASA EOS Data Centers. However, supporting documentation and other artifacts from these missions are also critical to the long-term studies of our planet’s climate. To be successful in this goal, we need to shift our focus from on-line Data Lifecycle Management, to the lifecycle management of information and what enables knowledge into the future. A break in that chain, like having incomplete documentation, can be deleterious for legacy missions especially when original investigations or people most familiar with the data have long moved on or are no longer accessible.

The Need for Documentation Preservation at NASA GES DISC

Gone is NASA’s not legislatively mandated to preserve data permanently, unlike agencies like USGS, NOAA and NARA, we have a challenge to develop a low cost solution that meets our data center needs and that of our users. The Goddard Earth Sciences Data and Information Services Center (GES DISC) has implemented a Repository System to facilitate the long-term archiving of documentation artifacts and other associated digital content. The GES-DISC designed this system based on Fedora Commons, an open-source repository management software, for cost savings and flexibility.

The first mission to utilize the GES-DISC Repository System was the High Resolution Dynamics Limb Sounder (HIRLDS) on the Aura spacecraft. Data and documentation from the Upper Atmosphere Research Satellite (UARS) and the Total Ozone Mapping Spectrometer (TOMS) and Nimbus have also been added. The GES-DISC is negotiating the transfer of data preservation items from the current Microwave Limb Sounder (MLS) on Aura, and the Atmospheric Infrared Sounder (AIRS) missions before they end.

Fedora Commons Interface

The GES-DISC used Fedora Commons, an open-source repository management system that is used in many universities, research centers, and libraries. It comes with a simple web-based GUI interface which provides easy administration of the system. The GUI also allows one to enter objects or data streams (these can be of any type document, image, source code, binary data, etc.) into the system. The system uses XMI to manage the objects. The GES-DISC has also developed a command line script to allow batch ingest of objects into the Fedora Repository.

Public Access of Preservation Documents

External users access the publicly available documents by visiting the mission specific documentation page for that instrument. The Fedora repository system is at the backend and makes access to the linked documents possible. Note that restricted objects (ITAR, proprietary, or software) are not accessible through the public interface. These missions are now public:

HIRLDS
TOMS
UARS

http://disc.csi.pds.nasa.gov/Aura/additional/documentation/hirdls-preservation-documents

GEDISC Preservation Implementation

An open source solution for document preservation to enable information generation from data for future generations of researchers

Fall 2014 AGU Meeting, Poster No. IN53C-3816

NASA/Goddard Earth Sciences Data and Information Services Center (GES DISC)

Irina Gerasimov1,2, Ed Esfandiari1,2, Michael Berganski1,2, Jennifer Wei1,2
1NASAGoddard Space Flight Center, 2ADNET Systems Inc.