

Harnessing NASA Goddard's Grey Literature: The Power of a Repository Framework

Nikkia Anderson, Gail Hodge, and Andrea Japzon

Information International Associates, Inc., NASA Goddard Space Flight Center, United States

Abstract

The NASA Goddard Library collaborated with several projects on-Center to create a framework for the development of web-accessible repositories of grey literature. Tools and methods for collaboration were developed through a series of prototypes with a variety of Goddard projects based on the Library's Digital Asset System (DAS), a repository to describe and provide access to project information including images, videos, web sites, and technical reports.

Metadata is a key component of this framework. The Goddard Core Metadata Element Set, an extension of the Dublin Core, is used to describe these resources. Additional elements include project name, project code and instrument name. A taxonomy of controlled subject terms has been developed which can also be extended and tailored for each project. This provides consistent searching across the DAS, while at the same time enhancing the search precision for each project collection when it is accessed as a separate collection in the DAS or through its own web site.

To-date, the Library has used the framework with a number of different projects. A commemorative CD with key documents, web sites and oral histories by the project managers was developed for the Hitchhiker Project as it was being disbanded. Metadata from the Swift project library was transformed and made more accessible with pointers to documents in the Swift project library. The Landsat Legacy Project, a joint project with the Landsat Program Science Office, the US Geological Survey, and the NASA History Office, is creating an archive of essential technical-, policy-, and science-related documentation. To gather significant documentation from the over 35-year-multi-agency history of Landsat, the Library added components for external submission of documents for potential inclusion in the new repository and for scanning paper documents. In addition, video histories are being captured from Landsat veterans.

Through these projects, the NASA Goddard Library has developed a methodology for collaborating with different kinds of projects. It involves procedures for analyzing the needs of a project and determining how the metadata, taxonomy, and interface might need to change, while remaining consistent with the DAS framework for cross-repository access.

Ultimately, the Library plans to extend the effort to other projects. (At any given point in time, Goddard has more than 30 projects in various stages of completion.) As part of Goddard's knowledge management activities, the DAS provides a framework for sharing grey literature that would otherwise be scattered across independent project libraries. Benefits include the ability to more quickly find and reuse information to decrease project costs, enhance safety and promote innovation.

Developing the Framework:

In 2001, the NASA Goddard Library began its efforts to develop a center-wide metadata repository of technical and scientific information related to the mission of the NASA Goddard Space Flight Center (GSFC). Through those efforts the Library has examined various approaches to making information accessible and available for re-use. An in-house prototype infrastructure for creating a combined metadata repository to describe and provide access to project information including images, videos, web sites, and technical reports was developed and called the Digital Asset System (DAS). The DAS was developed against the Goddard Core Metadata Element Set, based on qualified Dublin Core [1] with extensions. The Goddard Core Metadata Set was specifically developed to provide better discovery and evaluation in the Goddard context of project management [2], i.e., helping the user find the document, evaluate its usefulness and locate it whether in paper or in digital form. The Goddard Core was developed through the work of the Goddard Metadata Review Group (MRG), which consist of internal and external metadata experts with the primary focus on metadata for project documentation and related objects. The MRG supports the development of a metadata framework for digital archiving, preservation, and access across objects and subjects to advance knowledge management efforts [3].

To expand the elements of findability, the Goddard Library incorporated the NASA Taxonomy and EOS (Earth Observatory System) Taxonomy as controlled discipline terminologies for the subject element of the Goddard Core. The NASA Taxonomy was developed through a NASA Headquarters contract to the Jet Propulsion Laboratory (JPL), another NASA center located in California, in an effort to provide a strategic search mechanism across the NASA centers, thereby building a knowledge base to assist in browsing and navigation through large collections of disparate information objects [4]. The EOS Taxonomy was

developed at NASA GSFC to help support and serve the EOS Program Office for classifying information on a Web-portal [5]. In our efforts to increase the search precision over the years, while working with specific projects to gain access to important documents which no one could find, we developed project specific controlled vocabularies to bring those once grey documents to life.

Projects and Project Libraries (Home of the Grey Literature)

NASA GSFC has at least 40 active missions in various life cycle phases, at least 10 missions accomplished, and at least 15 missions planned to orbit. During these missions, NASA GSFC produces a tremendous amount of information. Some of these missions have project libraries that keep track of the information that pertains to the unique engineering and project management tasks involved in designing, implementing, launching, and maintaining the spacecraft [5]. There are some missions that don't have project libraries and the information is scattered about in various hands of the scientists, engineers, and project managers involved in the missions with little or no likelihood of being available for further use.

Through the Goddard Library's knowledge management initiatives, the Library became the face of resource discovery and information re-use in the project science and engineering arena at Goddard.

The Library was already known for providing assistance in finding information to suit the needs of its users, which are primarily the 9,000 employees (civil servants and contractors) on Center. Now with the knowledge management role, the Library was able to use other resources such as the Goddard Core Metadata Element Set and the DAS to help the users search information consistently across various project resources and not have to familiarize themselves with each individual project search system when available.

Starting Simple: The Hitchhiker CD

The venture into bringing grey information to life began with the Hitchhiker Project. The Hitchhiker Project began in 1984 and launched in 1986 [6]. In 2001, as the Hitchhiker mission was being hibernated, the Chief Knowledge Management Officer asked the Library to compile the information of the Small Shuttle Payloads Project to share with the project team and others interested in the project. Unlike other project collaborations that came after Hitchhiker, the Library didn't get a chance to create a sub-system of the DAS for Hitchhiker; instead we created a commemorative CD with web sites, images, key project documentation, and oral histories to serve as a finding aid for key information pertaining to the mission (see Figure 1).

Hitchhiker helped us see what was needed to develop sub-systems of the repository. Outlining the project scope was definitely a big need. When working on Hitchhiker, the Library was often at a loss, because the project team was no longer available to give the background of the mission and lessons learned. The Library was able to interview a few of the scientists and capture those interviews which were documented and made accessible on the resulting CD.

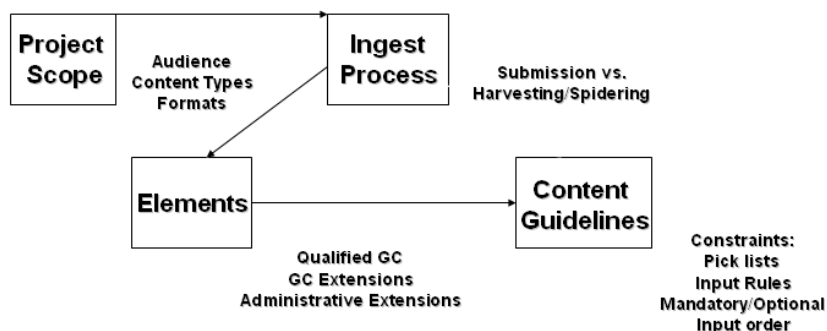


Figure 1: DAS Framework

Harvesting and Mapping a Project Library’s Metadata: The SWIFT Metadata Repository

The next effort at bringing the grey literature to light involved the Swift Metadata Repository and modifying the infrastructure of various project resources such as Goddard Project Directory, Goddard Project Libraries Database, and the IMAGES collection developed by the Goddard Library outside the DAS. During this time the framework went from a “what if” to a formalized way of searching and providing access to grey literature. The Swift mission launched in 2002 and is still in orbit with a focus on gamma ray burst. The Swift Metadata Repository was developed as a sub-system of the DAS with images of the Swift launch, web sites, videos, and pointers to project documents. The Swift project library uses a Centralized Configuration Management System (CCMS) for submission of documents, but the drawback is that the CCMS does not have a user-friendly search interface. In order to get information that you are looking for you must know as much as you can about that document to guarantee findability. So the Library took on the challenge of providing pointers with key information that would improve the access to the documents from the CCMS.

Swift Metadata Elements		GC Metadata Elements		
Number		Identifier.Original		
Payload		Subject.MissionProject/Subject.Instrument		
Responsible Person/Organization		Creator.Employee/Creator.Organization		
Title		Title		
Alt Number				
WBS Number				
Sub System		Subject.Instrument		
Revision/CH/DCN				
Submitted Date				
Approved Date		Date.Available		
Approver				
Remarks				

Figure 2: Swift Metadata Mapping to the Goddard Core

The Library mapped the metadata of the CCMS to the Goddard Core Metadata Element Set (see Figure 2) and incorporated other metadata elements required for the new content types. The Library developed a controlled vocabulary of terms specific to the Swift mission and customized some of the elements to project-specific controlled vocabularies. The user can simultaneously search across various content types using a single search box. Searching the metadata stored in the DAS retrieves the document which can be inserted into the CCMS search box to retrieve the document of choice. This system helps users discover the information resources of the Swift project through a single interface rather than separately searching for images, videos and documents.

Adding Ingest Options and Other Content Types: The Landsat Legacy Project

In 2004, the Goddard Library was contacted by the Landsat Program Science Office to help them create a Landsat document archive. The type of documentation sought for the Landsat archive includes policy-, technical-, and science-related materials with an emphasis on internal technical papers (or grey-literature) of both the government and the aerospace industry. The project has been dubbed the Landsat Legacy [7]. The Landsat Legacy team consists of a joint collaboration with the Landsat Program Science Office, the US Geological Survey, and the NASA Goddard Library [8]. The Landsat Program has been in existence since 1972 when Landsat 1 was launched. Currently the LPSO is on its 7th Landsat mission (Landsat 7). The DAS framework was incorporated to prototype this sub-system just as the others described above but with a few enhancements.

The original scope of this project was to include the policy-, technical-, and science-related documents but with further discussion regarding the possible deepness of the collection the idea struck of having video histories. The Library took on the role of recording live interviews of veterans of the Landsat Program over the last 35 years. By including the video histories to the scope of the project, the Library still continued to follow the process of the framework, which included selecting appropriate elements from the Goddard Core for the infrastructure and developing what is now the largest controlled vocabulary that we’ve created to date. The controlled vocabulary took on many versions before we got close to finalizing but with project-specific controlled vocabulary there is always room to expand. Drop down menus of controlled vocabulary for instrument, audience, and mission project were also added.

To gather significant documentation from over 35 years of the multi-agency history of Landsat, the ingest process of the framework was expanded to include additional components. The Library added components for external submission of documents for potential inclusion in the new repository and for scanning paper documents. The external submission of documents was set up to allow donors to submit documents through a web-interface that they may have stowed away in old filing cabinets or in boxes in their garage or home offices. Although the system is web-based the submission could be either electronic or paper documents, which allowed the Library to advance the DAS to include scanning of project documentation. The scanning process required procedures to be developed to ensure consistency and quality regardless of who did the scanning.

Since 2004 the Library has continued to make further advancements in the creation of the sub-system repository for the Landsat Legacy Project. Like the other DAS projects, the Legacy System will also simultaneously search across web sites, documents, and video histories. There may be efforts in the future to expand the project scope to provide access to the Landsat data archive for scientific images.

Conclusion – The Power of the Repository

Through these projects, the NASA Goddard Library has developed a methodology for collaborating with different kinds of projects. It involves procedures for analyzing the needs of a project and determining how the metadata, taxonomy, and interface might need to change, while remaining consistent with the DAS framework for cross-repository access. The framework allows for expanding in all areas as described through each project prototype. Through a series of challenging projects, the Library has improved the DAS and expanded the Goddard Core. The DAS and the Goddard Core have been presented as accomplishments through special presentations to the NASA Web Managers Group and the Metadata Review Group [9]. Through its efforts to harness the grey literature of NASA Goddard through a repository, the Library continues to provide consistent access, to share information across projects and to advance Goddard's knowledge management initiatives.

Acknowledgements

This work was conducted under NASA Goddard Contract NAS5-01161. The authors gratefully acknowledge the support and encouragement of Robin Dixon of the NASA Goddard Space Flight Center Library.

Works Cited

- [1] Dublin Core Metadata Initiative. "Dublin Core Metadata Element Set 1.1." [Online]. Available: <<http://www.dublincore.org/documents/dces/>> [November 6, 2006]
- [2] Senserini, A., Allen, R., et al "Archiving and Accessing Web Pages: The Goddard Library Web Capture Project." D-Lib Magazine. [Online] Available <<http://www.dlib.org/dlib/november04/hodge/11hodge.html>> [November 2004].
- [3] Metadata Review Group [Online]. Available: <<http://library.gsfc.nasa.gov/mrg>> [March 3, 2006]
- [4] Busch, J. (2004). "Taxonomy Use Examples." [Online] Available: <http://library.gsfc.nasa.gov/mrg/pdf/NASA_Taxonomy.pdf> [October 28, 2004]
- [5] Hodge, G., T. Templeton, et al "A Metadata Element Set for Project Documentation." Science & Technology Libraries, Vol. 25(4) [2005]
- [6] NASA Facts Sheet. "The Hitchhiker Project." [Online]. Available: <http://www.gsfc.nasa.gov/gsfc/service/gallery/fact_sheets/general/hitchh.htm> [November 13, 2006]
- [7] Rocchio, L., G. Hodge, T. Arvidson, D. Williams, and J. Irons. The Landsat Legacy: Tracking down three decades of knowledge. In: Pecora 16 "Global Priorities in Land Remote Sensing" Proceedings. American Society for Photogrammetry and Remote Sensing, Sioux Falls, SD. [2005]
- [8] Landsat Legacy Document Registry. [Online]. Available: <<http://library.gsfc.nasa.gov/landsat>> [July 25, 2005]
- [9] Anderson, N. & G. Hodge. "The Goddard Core: A Metadata Framework for Web Site and Beyond - Presentation at Web ManagerViTS." [Online]. Available: <http://library.gsfc.nasa.gov/mrg/pdf/GC_WebManagersViTS_07242006.pdf> [July 24, 2006]