Wallops Island Balloon Technology:
Can't see the Repository for the Documents

Andrea Japzon and Nikkia Anderson
Information International Associates, Inc.
NASA Goddard Space Flight Center

Abstract
Since the Wallop's Balloon Technology documents repository began approximately 9 years ago, the Goddard Library has become increasingly involved in developing digital archiving capabilities. The Library developed the Digital Archiving System (DAS) which is a prototype infrastructure for creating a combined metadata repository that allows metadata for heterogeneous digital objects to be searched with a single search mechanism and presented in a single results page. With this, the opportunity has been presented to expand the usability of the print repository. The Balloon Technology documents relate only to the specific subject of construction of scientific balloons and at the current time number over 4,300. The documents exist primarily in paper format and are organized according to the accession number. The project is currently at a crossroads where decisions will have to be made regarding the reorganization of the database from many different perspectives. An assessment of the project was conducted to determine future direction. An assessment survey was created using the Knowledge Management Assessment Tool (KMAT) from the American Productivity & Quality Center and from the recommendations that The Scholarly Publishing & Academic Resources Coalition (SPARC) put forth in "The Case for Institutional Repositories: A SPARC Position Paper. Survey participants agreed to move forward with project by scanning the documents, mapping existing database records to the current metadata elements, seeking copyright permissions, and forming a joint committee of balloon scientists and librarians. All have agreed on the importance of digitizing this collection to the balloon science community. Further, it was agreed that once complete, the addition of the balloon documents to the DAS (an institutional repository) could serve as a model for other NASA and/or government projects trying to organize, collect and preserve specialized knowledge that manifests largely in grey literature.

History of the Library and the Database

The Balloon Technology Library and Database was created to provide a single repository for scientific ballooning literature and data. Grey and commercially published literatures comprise the collection of technical reports, working papers, proceedings, and journal articles. The primary impetus for creating the library and database was the fact that much of the literature relating to balloon technology is grey literature and therefore not easily found, let alone found in one place. The Balloon Technology documents, currently numbering over 4,300, relate specifically to the construction of scientific balloons. The subject matter covered includes balloon theory, balloon physics, design, operations, performance, facilities, testing, materials, fabrication, quality control, failure analysis and history [1].

The Balloon Technology collection project started in November of 1993 continues to exist today with the sustained involvement of Wallops Flight Facility (WFF) balloon scientists and Goddard Space Flight Center (GSFC) librarians. WFF is a part of GSFC but located three hours south of Greenbelt, Maryland in Wallops Island, Virginia. The WFF mission includes the performance of research in the areas of sub-orbital and small orbital payloads. Balloon science experiments are a significant part of this mission. From the beginning of the project, the intention of the Balloon Program Office (BPO) at WFF was to start a library that would become the World's Archival Center for Scientific Balloon Technology [2]. To that end, project participants set out to and accomplished the major goals of collecting to the greatest number possible of existing balloon technology documents and creating a searchable database. With these accomplishments in hand, now is the time to increase the availability of this work to balloon science researchers.

A great debt is owed to Jim Winker, balloon scientist and de facto librarian, for performing the yeoman's duty in reaching the goals set out by the BPO. In the field of librarianship, perhaps the greatest value placed on information is its authority. In the creation of this collection, Jim Winker's life experiences provide that authority. He possesses extensive historical and technical knowledge of scientific ballooning. His knowledge has developed through "a long continuous interest and participation in the field of scientific ballooning and interaction with the many present and past commercial and government organizations" [3]. The knowledge and experience of individuals within a specialized field is what sustains the use and value of grey literature within the field. Jim Winker's large-scale interaction with balloon literature clearly exemplifies this.

Mr. Winker visited over 60 sites in his effort to find and select documents for the library [4]. He started locally with his own collection, the Raven Industry library, private collections in Sioux Falls, South Dakota and the holdings of Wallops. His search expanded to include the information repositories of the National Technical Information Services (NTIS), Defense Documentation Center (DDC), Association of Balloon and Airship Constructors (ABAC), and American Institute of Aeronautics and Astronautics (AIAA).

He explored organizations that use balloons currently or have used them in the past. These organizations included research centers, manufactures, scientists, research sponsors, and operations centers. In his effort to be comprehensive, his search included an "everything else" category as well. He evaluated the resources of libraries, museums, individuals, published works, and international resources [1].

With the care of an archivist, Mr. Winker researched preservation quality paper, durable binding, and copyright implications [5]. While he realized the selection and preservation of the documents was of fundamental importance, he also realized that "the database may well be the most important part of the project. Without it, the library would have a far more limited benefit" [1]. The database would promote the existence of the collection and ultimately provide access to the documents. Working with Janet Ormes, GSFC Library Head, Mr. Winker, selected software to support the database and created guidelines for cataloging the documents. The *COSATI Standard for Descriptive Cataloging of Government Scientific and Technical Reports* was relied upon for the creation of the rules of entry [6]. For each document, an entry was created for the database using a variety of descriptors intended to aid in the process of information discovery.

The askSam database was chosen in 1994 and began production with the database successfully organized and defined [7]. At this same time, database search methods were explored and decided upon. Progressive for the year, both controlled vocabulary and natural language searching were made possible [8]. In the ten years that the documents were collected and that the database grew in significance for the ballooning science community, technology was developing that would further enhance the work of Mr. Winker.

Current Status of the Documents

Since the Wallops' Balloon Technology Library began, the Goddard Library has become increasingly involved in developing digital archiving capabilities. The Library developed the Digital Archiving System (DAS), a prototypical infrastructure for creating a combined metadata repository allowing for heterogeneous digital objects to be searched with a single search mechanism and presented in a single results page. The DAS is an institutional repository of scientific and technical information including selected Goddard web sites, videos, images and documents. With this, the opportunity has been presented to expand the usability and accessibility of the balloon technology library.

The documents exist primarily in paper format and are organized according to the accession number. The documents are housed as a separate collection at the Wallops Island Technical Library. The database is accessible to all from the library's website. NASA researchers can gain access to the documents by either using the library at Wallops or having the documents scanned and emailed to them. Researchers outside of NASA can make requests from Wallops with some restrictions placed on their access.

Since January of 2004, some balloon technology documents have been scanned as a pilot project into a MySQL database as PDF files. The askSam records as contained in a WAIS database link to the PDF files; however, the DAS which the MySQL database supports does not relate to the askSam database. The askSam software is no longer being developed. It has limited capabilities for interacting with the World Wide Web and the searching environment. Currently we have two separate search interfaces, one is a meta search in WAIS and the other Autonomy which indexes the documents and has full text search capabilities [9,10]. WAIS and Autonomy can't be configured to work together to provide meta and full text searching. The DAS can do both by using MySQL and Lucene. Not to mention, the DAS allows you to search, archive, and preserve documents for future use. By incorporating the balloon technology documents into the DAS, the balloon science community will have increased visibility and access through meta, full-text searching and retrieval options to include the full-text.

The project is currently at a crossroads where decisions will have to be made regarding the incorporation of the database into the DAS and the accessibility of the documents on many different levels and perspectives. The authors of this paper turned to the community involved in the creation, the delivery and the use of the library and database to determine the future direction of the project.

Knowledge Management Assessment

An assessment survey based on the Knowledge Management Assessment Tool (KMAT) from the American Productivity & Quality Center was administered to nine individuals, representing different perspectives on the project [11]. Further, these same individuals assessed the project based on the recommendations that The Scholarly Publishing & Academic Resources Coalition (SPARC) put forth in "The Case for Institutional Repositories: A SPARC Position Paper [12]. See Appendix A for the complete survey assessment. A meeting was held on August 12, 2004 with survey participants and Balloon Program Office (BPO) representatives in attendance. Issues brought forth by the assessment created the agenda for the meeting.

Given that the work of the Balloon Technology collection is a small slice of a greater organizational structure, survey participants were ask to be cognizant of this while responding to the KMAT as knowledge management is the function of an organization. The authors felt that the current

status of the project presented an instance of the process of knowledge management in play. Mr. Winker is retiring for the second time in January 2005 and will no longer be working on the project. As stated earlier, he has performed the vast majority of document collection and cataloging. The authors wanted to examine how well the knowledge that Mr. Winker possesses is being transferred to the individuals remaining on the project.

The greatest benefit of the knowledge management assessment indicated an important difference in perspective among survey participants. Two perspectives emerged: content and infrastructure. Individuals with a content perspective included contributors and users of the information contained within the collection. Individuals with an infrastructure perspective included systems, web, library, and archival staff that primarily work to make the collection accessible to users of the collection.

At the assessment meeting, both perspectives were defined and the importance of each established. Certain points of confusion regarding current communication between all project participants were made clear. Traditionally, the individual who creates information and/or collects the information is not the same individual that catalogs, stores, and distributes the information. For ten years, Mr. Winker was performing most tasks related to the collecting, cataloging, and searching of the documents, not to mention creating literature as well. Enter the possibility of the DAS and open access archives and the flow of knowledge from old processes to new processes does not transfer without some adjustment. The second part of the assessment devised from the SPARC paper complimented the first as it helped the discussion participants describe the processes under consideration for change.

Can't see the Repository for the Documents

For the purpose of this paper, the definition of an institutional repository as put forth by the Scholarly Publishing & Academic Resources Coalition (SPARC) will be used. Institutional repositories are digital collections that capture and preserve the intellectual output of a particular community. Repositories are intended to expand access and retain control over the scholarship produced by that community. Further, the repository has the potential to contribute to the community or the institution it is a part of by providing tangible indicators of the community's quality through the demonstration of the scientific value of its research activities [12].

Being a part of the DAS can in and of itself increase the visibility, status and public value of the balloon science community. Money, time, and technology aside, the documents themselves create the biggest obstacle to creating an institutional repository at Goddard as put forth by SPARC. The SPARC paper further defines an institutional repository as "open and interoperable." To promote interoperability and open access, institutional repositories provide access to the broader research community through no or low barrier access. Either by providing a search mechanism with indexing or by maintaining and exposing metadata to be harvested by other institutions, interoperability is gained among institutions [12].

While the BPO is interested in sharing the database of records in the manner described above, the office is not interested in sharing the documents in a global capacity. The proprietary nature of some of the documents and the competitive environment in which they were created prohibit the collection of documents as a whole from being accessed outside NASA IP ranges. Small Business Innovation Research (SBIR) documents are an example as they are restricted for five years after their release. The BPO representatives want to keep the collection primarily for balloon scientist with a relationship to NASA research, hence limiting access to greater research community. It is extremely likely that other Goddard projects will have similar concerns or restrictions regarding their documents. Fortunately, the technology exists to allow different levels of access to different types of information allowing Goddard to interoperate fully in at the metadata level.

If the database records are to be open and interoperable, the existing data fields from askSam will have to be mapped to the metadata elements of the Goddard Core. The Goddard Core is a metadata element set and is employed in the DAS single search mechanism. Metadata is information that describes a digital object; like a library catalog card for a digital object. The Goddard Core Metadata set contains 24 elements that describe project-related objects of interest to GSFC. It extends ISO 15836 (Dublin Core) Standard for Descriptive Metadata for Electronic Resources while using Open Archives Initiative standard protocols that can interoperate with other systems.

When Mr. Winker was devising his catalog entries, he was working in a contained environment. There was no need to consider how the descriptive fields he employed related to a larger information structure. The technology did not exist for interoperability to be an issue. The Goddard Core establishes guidelines for organizing documents and other objects in a way that is both meaningful and accessible across GSFC, within NASA and in the open access environment.

The mapping will involve more than a programming solution as the fields used in askSam do not match directly or in some cases not at all to the Goddard Core elements. Without ISO standards in place, definitions of descriptors were not nearly as controlled. The Goddard Core allows for project profiles so if need be the 24 elements can be expanded to accommodate the particularities of a document or object collection. The mapping of askSam fields to Goddard Core is a key area for the collaboration from both content and infrastructure perspectives.

Copyright presents another problem in bringing an already existing print library and database into the realm of the online repository. To date, there are 4,330 records in the database of which 847 are indicated as copyrighted. That is roughly 20%. A copyright campaign will need to be launched on behalf of the repository under the auspice of the BPO. This will take a significant amount of time and agreement is not guaranteed. It was agreed that copyrighted documents will begin to populate the DAS and as permissions are gained documents will be added with the understanding that all documents are not likely to be included in the online repository.

As mentioned previously, Mr. Winker is retiring. His knowledge is the current certification and collection process. While encouraged by the ease of collecting new submissions to the collection through online technologies, the users of the collection and the BPO representatives were concerned about the quality that self-selection produces as well as the possibility for submitting items out of scope. Within the balloon science community, there is considerable interest in expanding the collection to include literature related to the experiments conducted using the balloons. Further, online submissions would not replace the need to produce an archival paper copy for the balloon technology library. The meeting participants decided that an official joint committee of Goddard librarians and BPO staff should be sanctioned to establish a certification process and to address any further decisions that need be made regarding the repository. Additionally, the joint committee will facilitate and encourage communication between infrastructure and content perspectives.

Conclusion

As the survey participants sat in the room where the Balloon Technology library is housed, a collective realization fell over the group at the conclusion of our assessment discussion. The ways in which people seek and interact with information are dynamic and online. The print collection resting on the shelves seemed anachronistic given our discussion.   The BPO and the Goddard Library agreed to move forward with scanning the documents, mapping the records, seeking copyright permissions, and forming a joint committee. All have agreed on the importance of digitizing this collection to the balloon science community. Further, it was agreed that once complete, the addition of the balloon documents to the DAS institutional repository could serve as a model for other NASA and/or government projects trying to organize, collect and preserve specialized knowledge that manifests largely in grey literature.

Currently, funds are being pursued to make the transition from a library and a database to an institutional repository a reality at Goddard. Given the nature of the types of information produced at Goddard, any institutional repository created at Goddard or a like environment will not be able to participate fully in open archive initiatives. In our case, the DAS works as an institutional repository for Goddard and to a lesser degree beyond the walls of NASA IP addresses: metadata is acceptable for web harvesting but most documents would not be freely available.

The DAS is still in beta testing and contains over 90,700 web pages, 900 images, and 400 videos. The Balloon Technology document collection will be the first document collection to be included in the DAS. The choice of this collection has proven fortunate in that it exposed many of the issues we would likely encounter when considering the addition of other collections.

Author Information

Ms. Andrea C Japzon
Access & Navigation Services Team Leader
Information International Associates, Inc.
NASA Goddard Space Flight Center
Code 292, Bldg 21, Rm L112
Greenbelt, MD 20771
USA

ajc1170@aol.com
http://www.gsfc.nasa.gov

Works Cited

(1) Winker, J. A. (1994). A Balloon Technology Library Database. Paper from the *32nd Aerospace Sciences Meeting & Exhibit.* Reno, NV, January 10-13, 1994.

(2) Winker, J. A. (1993). Balloon Technology Library Project Monthly Progress Report No. 1. December 3, 1993.

(3) Winker, J. A. (1993). Balloon Technology Library: Discussion topics for initial project meeting of 23 November 1993. November 18, 1993.

(4) Winker, J. A. (1998). Balloon Technology Library Project Monthly Progress Report No. 45. January 2, 1998.

(5) Winker, J. A. (1994). Balloon Technology Library Project Monthly Progress Report No. 3. February 1, 1994.

(6) Suggested Guidelines. (1994). Condensed from *Guidelines for Descriptive Cataloging of Reports: a Revision of COSATI Standard for Descriptive Cataloging of Government Scientific and Technical Reports.*

(7) Winker, J. A. (1994). Balloon Technology Library Project Monthly Progress Report No. 5. April 2, 1994.

(8) Winker, J. A. (1994). Balloon Technology Library Project Monthly Progress Report No. 6. May 4, 1994.

(9) *Balloon Technology Database*. (2004). <http://library.gsfc.nasa.gov/Databases/Balloon/balloon1.html> (October 26, 2004).

(10) Search Goddard Library. (2004) <http://library.gsfc.nasa.gov/SiteSearch/premium/premiumquery.html> (October 26, 2004).

(11) American Productivity & Quality Center. (2001) *The Knowledge Management Assessment Tool (KMAT)*. <http://www.kwork.org/White_Papers/KMAT_BOK_DOC.pdf> (October 26, 2004).

(12) Crow, Raym. (2002). The Case for Institutional Repositories: A SPARC Position Paper. <http://www.arl.org/sparc/IR/ir.html> (October 26, 2004).

Appendix A

Part I

The KMAT was used as presented at http://www.kwork.org/White_Papers/KMAT_BOK_DOC.pdf with the following questions removed: P3, L2, M1, and M3. Reason being the project is produced by a government organization and therefore will not be sold or marketed for profit.

Part II

1. What purpose does an Institutional Repository serve?
2. Policy consideration:
   a. What is the current copyright policy? Does it need to be improved?
   b. How are restricted documents handled? Could this be done differently?
   c. With regards to accessibility, do we need a policy that differentiates between internal and external customers?
   d. What is the current certification process, the process that assures the quality of the documents added to the repository? How will this change once Mr. Winker leaves the project?
   e. When you consider the future of the repository, will we need a formalized accession policy?
3. Does the repository embody the institutional quality of
   a. The Balloon Technology Program   YES ☐   NO ☐
   b. NASA   YES ☐   NO ☐
   c. If No to either "a" or "b", please explain what needs to change for this to be so?
4. Does the repository have formal or official recognition in the Balloon Technology Community? Please explain why or why not.
5. On a scale of 1 to 5, please indicate how important you think it is that the Balloon Document Repository reflects the following. 1 being not at all important and 5 being very important.
   a. Institutionally Offered _____
   b. Scholarly _____
   c. Cumulative and Perpetual _____
   d. Open & Interoperable _____
6. Please indicate by checking which of the following you think should be reflected in the metadata? Jim Winker's Metadata Elements have been mapped to the corresponding Element in the Goddard Core.

Jim Winker's Metadata Elements Used	Goddard Core Metadata Elements Used	Please Check X
Access Number	Identifier.Persistent (Auto)	
Title	Title	
Author	Creator.Employee	
Responsible Organization	Creator.Organization	
Funding Organization	Contributor.Organization	
	Creator.RecordCreator (Default)	
Date	Date	
Notes		
	Date.Current	
	Date.Created (Auto)	
	Date.RecordCreated (Auto)	
Report Number	ID	
Contract Number	Contributor.Contract	
Description		
Subject Terms	Subject.Uncontrolled	
Content	Description	
Available From		
	Contributor.Code	
	Type (Default)	
	Content.Type	
	Format (Auto)	
	Subject.Discipline	
	Identifier.URL (Auto)	
	Language (Default)	
	Rights (Default)	
	Audience (Default)	

7. Please describe any other elements that you think should be added?
8. Please feel free to make any additional comments regarding the Balloon Technology Documents?