Early, C., Japzon, A., & Endres, S. (2006). Creating a knowledge base: Analyzing a veteran reference librarian's brain. In Lankes, R. D. (Ed.), *The virtual reference desk: Creating a reference future* (pp. 155-68). London: Facet Publishing.

Chapter 8

Creating a Knowledge Base: Analyzing a Veteran Reference Librarian's Brain

Charles Early, Andrea Japzon, and Sarah Endres

OVERVIEW

Virtual reference transactions provided a solution to a knowledge-management problem at the NASA Goddard Space Flight Center Library. Online reference services, real-time and e-mail, allow for the development of a knowledge base. One hundred e-mail question-and-answer pairs were analyzed to reveal the steps taken and the sources used by a soon-to-retire librarian. A guide to the information sources and producers at Goddard was created. Given the dynamic nature of information at Goddard, a database built on the Apache, MySQL, PHP (AMP) open source platform was designed for the guide. The resource is now dynamic and can continue to grow with input from all Goddard's librarians.

INTRODUCTION

Goddard Space Flight Center is located in Greenbelt, Maryland, 10 miles northeast of Washington, D.C., and was established by the National Aeronautics and Space Administration (NASA) as its center for space research in 1959, soon after the agency was founded. It has become one of the world's leading research establishments, with the largest scientific staff of any of the NASA field centers. Nearly 6,000 scientists and engineers work at Goddard, in addition to well over a thousand contractors working for Goddard at nearby off-site locations. These people are the Goddard Library's main users.

Goddard's primary mission today is building, designing, and using scientific research satellites in Earth's orbit (along with the scientific instruments and observatories they carry). These include astronomical satellites such as the Hubble Space Telescope and earth observation satellites such as Landsat and the weather satellites that Goddard builds for the National Oceanic and Atmospheric Administration. In support of this primary mission, Goddard also maintains a network of satellites and ground stations for tracking and data communications, and several enormous digital repositories of data from the satellites. These activities all pose major engineering challenges.

Engineering, especially the cutting-edge, highly specialized sort of engineering that goes on at Goddard, relies heavily on unconventional, often fugitive, information resources. For example, these resources include technical reports, standards and specifications, engineering drawings, design guidelines, and lessons learned. Institutional (Goddard and NASA) information and knowledge resources are essential for Goddard engineers, but scattered and difficult to locate.

NASA's Scientific and Technical Information Program was a pioneer in the use of computers to manage information in the early 1960s (see www.sti.nasa.gov). Today it maintains a database of nearly 4 million citations to aerospace-related documents and makes a collection of over 1.2 million technical reports available in hard copy and electronically through the Center for Aerospace Information. Other NASA information systems include the online union catalog of the NASA Center libraries, a database of NASA and other standards and specifications, a database of NASA directives and regulations, a metasearch engine for the many NASA photo and multimedia collections, and still others. Additional information resources are scattered among the thousands of NASA and Goddard Web sites. Some important documents are still available only in hardcopy in undisclosed locations. Bibliographic control is excellent for the major NASA databases, and poor to nonexistent for everything else.

When Goddard begins a new project to design and build a satellite, an ad hoc work group is formed. The design and construction process generates hundreds or thousands of technical documents: plans and drawings, design reviews, modifications, test results, and so forth. These documents need to be managed while they are being used by the project, and a project library is set up. This library is intended only for use by the project group, and usually little or no effort is made to offer any sort of access to anyone else. Once the project is completed and the satellite is launched, these documents are warehoused and forgotten, and much valuable and reusable knowledge is forgotten with them. Some progress has been made in improving this situation. The Goddard engineering librarian, Early, maintains a directory of active project librarians and a mailing list for project librarians who will cooperate and share knowledge. There is a Goddard-wide database, the Centralized Configuration Management System, that holds about 10,000 documents from some Goddard projects, but it is incomplete and limited in scope.

VIRTUAL REFERENCE A SOLUTION TO A KNOWLEDGE-MANAGEMENT PROBLEM

Because some institutional information sources are extremely decentralized, poorly documented, and poorly archived, the Goddard Library has depended almost entirely on the experience and tacit knowledge base of Early, the senior librarian, who is soon to retire, to handle this class of reference requests. There was essentially no backup. A new solution for answering difficult reference questions was needed to provide a means for new or less experienced staff to answer questions without Early. First, the relevant and helpful knowledge possessed by Early needed to be identified. Once identified, it need be captured and made accessible at future points of need. While Goddard presents a particularly complex information environment, most libraries deal with the same problem on some level.

Libraries have always sought to make explicit, that is, to codify or articulate knowledge for the purpose of documentation, the tacit knowledge of reference librarians. "Tacit knowledge is intuitive and practice-based, which makes it both valuable and difficult to pass on to others" (Stover, 2004). Historically, reference departments have employed a hard-copy card file rolodex-style system for capturing and updating difficult to find information. The Ithaca (N.Y.) College Library and Multnomah County (Ore.) Public Libraries provide early examples of libraries employing online technologies to create a knowledge base for reference work. Ithaca populated a FileMaker Pro database with questions from the reference department card file and then expanded the knowledge base from there. Multnomah used e-mail to track and ticket the workflow of reference requests while using a database engine to control and organize question answering (Perez, 1999).

The reference librarians of New Brunswick (N.J.) Campus Libraries of Rutgers University created CKDB, the common knowledge database, as a means of sharing knowledge across campus libraries and to aid librarians in answering unfamiliar discipline-specific questions. (See http://ckdb.rutgers.edu to access the knowledge base.) CKDB is reported to have improved communication across campus libraries; however, the process of contributing to the database is not embedded, but voluntary. Therefore, stories about the steps taken to find answers to difficult reference questions are not routinely or systematically captured in the database (Jantz, 2001). Questions answered at a reference desk rather than via a virtual reference system are more difficult to archive as the process of answering a question is not automatically captured.

The Web-based Ready Reference Database (RRD) at San Diego (Calif.) State University provides a highly structured example of a reference knowledge base. The RRD project leader actively collects information from the reference staff through both formal and informal means. Then, the leader determines which aspects of the information provided are most critical and relevant to the information that already exists in the database. Through this process, the leader maintains the database with updates and cross-references (Stover, 2004). Thus, the RRD provides access to valuable information but not to steps taken to obtain the information.

Online reference services, real-time and e-mail, facilitate the development of knowledge bases. With virtual reference transactions, a tangible artifact is created with each exchange, which is usually stored automatically. Librarians can now archive particularly meaningful exchanges to be used again to answer future questions. With this, the potential exists to record the tacit knowledge of reference librarians making explicit their knowledge, which users and library staff can learn from. The transcripts from real-time sessions and e-mail correspondence provide a new vantage point for studying reference staff behavior, the research process, and resource usage. For example, the Digital Reference Education Initiative (DREI) provides access to both good and bad chat reference transactions for the purpose of teaching the reference interview through real interactions (see http://drei.syr.edu/item_list.cfm?NavJD=22).

OCLC's QuestionPoint is both a service and a product. In addition to being a tool for managing virtual reference exchanges, QuestionPoint is a searchable global knowledge base that is created by, and provided to, all its customers. The acquisition of 24/7, a chat reference service, by OCLC will no doubt expand the 1,000 libraries using the product and thereby increase the over 7,000 question-and-answer pairs in the database (OCLC, 2004). The knowledge base has the potential to save librarians time and to provide knowledge of new resources through the collective sharing of the knowledge and experience of many reference librarians.

The knowledge base is a tool of knowledge management in that it provides a means of generating, capturing, and sharing the knowledge contained within. The value of human and intellectual capitals has increased because of the current state of information economy (Hirsh and Dinkelacker, 2004). Given this, the United States government and the corporate sector have struggled with ways of keeping the knowledge of employees long after retirement (Liebowitz, 2002; Hoffman and Hanes, 2003). They are seeking ways to formalize or codify knowledge, that is shared or used in an informal manner (Liebowitz, 2002).

The knowledge of scientific and technical workers provides the most tangible examples of the problems that can arise when knowledge is not documented. Technology changes at a rapid pace, and infrastructures based on that technology often do not change as rapidly. In the case of NASA, the loss of engineers from the Apollo Era meant a loss of their knowledge (Hoffman and Hanes, 2003). The Electric Power Research Institute (EPRI) study of electric utilities workers found that manuals and procedures were not effective tools for eliciting tacit knowledge leaving much knowledge undocumented (Gross, Hanes, and Ayres, 2002).

As in the examples of library reference knowledge bases detailed above, most library knowledge-management systems/resources capture explicit knowledge: topical resources, location of resources, and the best possible resources for different query types. These systems do not typically succeed at capturing tacit knowledge: knowing how to find information, what is available, how to select relevant sources, and how to follow a path to the right information (Gandhi, 2004). The EPRI study recommends devising a plan to elicit, store, retrieve, and present knowledge when needed. It could be argued that virtual reference systems do just that for the knowledge of reference librarians. While Bill Katz's (2002) two-volume introduction to reference services provides a vital piece of instruction for reference librarians, the transcripts from virtual reference transactions have the potential to document the tacit knowledge of librarians, thereby detailing a fuller understanding.

DEVELOPMENT OF THE GUIDE

Like most reference departments, for years the Goddard Library has kept a log of reference questions. Questions answered in person, over the phone, and via e-mail are transferred to an Excel spreadsheet. In the process, the details of the transaction are greatly reduced. In March of 2004, a chat reference service was launched at the library. The service automatically generates a record of each reference transaction, thereby creating a knowledge base of questions and answers. This service, unlike e-mail reference, is not well used at Goddard. So Early and Japzon began to populate the knowledge base with entries from the

Excel question log in an effort to create a tool for answering questions. They found that this Q&A resource provided little orientation for librarians in this information environment. Where does one begin, and how does one get access to the right question to begin answering the current question? As mentioned previously, a significant amount of information at Goddard is produced from many different internal projects and is not centrally managed. The goal became to organize Early's knowledge and navigation of Goddard and NASA information rather than organizing the sources of information.

Early suggested emulating the Department of Defense's (DOD) *How to Get It Guide* (Doezema and Fox, 1998). This is a print volume updated at irregular intervals that is intended to aid individuals in accessing DOD information. Early and Japzon sat down together with pen and paper waiting for Early's knowledge about Goddard to surface in an orderly manner. This was not a productive effort. They found they needed some cues to elicit the knowledge from Early. They went back to Questions & Answers for the cues. Early randomly selected 100 previously asked questions from the outbox of his e-mail account. They tried again, this time with Endres, the cataloger/reference librarian, and began to describe and classify the information sources and the information producers at Goddard that were used to answer the questions.

After the review of approximately 70 questions, the point of diminishing returns was reached. It took much longer to analyze the first questions than later ones. The mention of a resource or an agency produced a mental association with others in Early's mind, and these resources were documented along with the steps he took to answer the questions. The result was 160 distinct entries for either information resources or producers related to Goddard.

The entries were added to an HTML document and were organized into four categories: NASA documents, technical reports, open literature, and images and multimedia. As the list of resources increased, it became more difficult to keep the HTML document organized and properly formatted. It was then decided to transfer the information to an AMP platform database, which was created by one of the library staff.[1]

Various decisions were made regarding the format of each record and the standards for entering information. We decided to separate the entry into several parts: title, URL, description, and keywords. The title and URL have a specific format to which they should adhere, and a style sheet was created to help people enter their information in the correct format. After discussing the benefits of a controlled vocabulary versus uncontrolled keywords, the team decided that uncontrolled keywords would be of greater use. The person entering the information in the database could type in any keyword he or she thinks a user might enter to find this information. Figure 8-1 illustrates the search results for the keyword search *standards*.

Once the database structure was created, it was necessary to decide who would be using the database and who would be updating it. Since the guide would have resources that are limited to Goddard employees, either because of license agreements or security restrictions, it was decided to keep the guide limited to users within the Goddard domain. However, the perceived primary users of the guide, and the only ones adding to

Figure 8-1
Search Results for the Keyword *Standards*

Office of Scientific and Technical Information (OSTI)	http://www.osti.gov/	Office of Scientific and Technical Information of the U.S. Departi Energy (DOE). The information clearinghouse for DOE and its predecessor agencies ERDA (Energy Research and Developme Administration, 1974-1977) and AEC...more
Office of System Safety and Mission Assurance Office of System Safety and Mission Assurance (OSSMA)	n/a	A major division of GSFC (Code 300). Formerly OFA. See Code Specifications. more
Orbital Information Group	http://oig1.gsfc.nasa.gov	A group within the Space Communications Program which mair database of detailed information about the orbits of all artificial s in Earth orbit. The database also has information about man-ma debris; i...more
Outgassing Handbook	http://outgassing.nasa.gov/	A searchable database information about gasses given off by va materials under vacuum conditions, derived from tests conducte GSFC from 1967 through the present. Available in print form as Outgassing Data for Selecting Spacecraft Materials ...more
Parts Information	n/a	Engineers often need to know technical specifications for parts considering using in a design (electronic components in particul Mechanical Engineering Library has a sitewide licence for acce: collection of vendor catalogs at h...more
Procurement-related Information	n/a	The NASA Aquisition Internet Service (http://nais.nasa.gov/) has extensive information resources for anyone involved in doing bus with NASA, including the NASA Procurement Reference Library includes the NASA Procurement Management System (...more

the database, are the library staff and the various project librarians located in different buildings on campus. To maintain quality control of the database, any entry created goes into a temporary database, where it is then reviewed by one of the reference librarians before being added into the permanent database. This allows the records to be edited for corrections. Figure 8-2 shows an example of a database entry.

Now that the database has been created, and the entries added, how does one find the correct entry? Three different ways of finding the correct entry were devised: searching, internal linking, and a show-all feature. The searching feature currently searches every field in the record. There are four kinds of searches to choose from: *substring* (which is the default), *AND*, *OR*, and *exact phrase*. For most purposes, the sub-string search is the most useful as it searches for any record that contains the string of characters somewhere in the text, even if the string is

Figure 8-2
Example of a Database Entry

NASA Handbook (NHB)

URL: *n/a*

NASA Handbook, a series of NASA directives that have been superseded first by NPG, NASA Procedures and Guidelines, and since December 12, 2003 by NPR, NASA Procedural Requirements. Likewise, the GHB (Goddard Handbook) series has become GPG, and similarly for other NASA centers. The series NMI (NASA Management Instruction) is now NPG (NASA Policy Guideline); similarly for GSFC and the other centers. The Cancelled Directives Report from NODIS has the history and current replacements (if any) of NHBs and NMIs, as well as more recent directives.

- NASA Handbook 5300.4 (NHB5300.4)
  A group of 15 directives related to quality assurance and workmanship. All parts of NHB5300.4 were cancelled around 1996. Some were first converted into NASA Assurance Standards (NAS) with the corresponding number for a brief period. (See http://www.hq.nasa.gov/office/codeq/doctree/docgenis.htm). Their replacements, if any, are listed below:
  - NHB 5300.4 (1A - 1): Deleted. Partially replaced by NASA STD 8729.1 and NSTS 5300.4 (1D - 2), Chapter 3.
  - NHB 5300.4 (1B): Replaced by NPD 8730.3 (superseded by NPD 1280.1) and the industry standard ANSI/ISO/ASQC-Q9001-2000 (an ISO 9000 standard).
  - NHB 5300.4 (1C): Replaced by NPD 8730.3 (superseded by NPD 1280.1) and the industry standard ANSI/ISO/ASQC-Q9002-1994 (an ISO 9000 standard).
  - NHB 5300.4 (1D - 2): Replaced by NSTS 5300.4 (1D - 2).
  - NHB 5300.4 (1E): Deleted.
  - NHB 5300.4 (1F): Deleted.
  - NHB 5300.4 (1G): Deleted.
  - NHB 5300.4 (2B - 3): Replaced by NPG 8735.2 (now NPR 8735.2).

within a longer word. Internal linking was also added, so if a user found mention in a record of a hyperlinked resource of interest, a click on the word links to the entry about that resource. The last method of record retrieval is just to comb through all the records using the *show all* feature.

IMPLEMENTING THE DATABASE

The entries from the HTML guide were copied into the database and keywords added to each record. Then, two of the reference librarians did a quality-control check of the database: editing records, removing duplicate entries, merging other records together, and adding internal linking (which did not transfer from the HTML document). The librarians then tested the database for ease of use and came up with a few corrections, which were then implemented. A cheat sheet was created to help users search the database and enter new records. A training session was held for the library staff and project librarians in which the different features were demonstrated and any questions answered. The database is not a finished product, so users are encouraged to provide feedback via e-mail, phone, or an online survey.

Plans are progressing to enhance the database. The librarians intend to add more fields to aid in finding the documents in the future and to improve searching capabilities. The number of keywords will be increased, and the searching options will be expanded so specific fields can be searched and limits can be placed. Show-all functions will be created for the original four categories, so a users can essentially browse the original ordering of the entries, which will allow them to catch records that they may miss with the search.

BENEFITS

The guide is already a great benefit to the library staff and, with future improvements, will become even easier to use. On February 15, 2005, the guide in the database format was made available to both Goddard Library librarians and the project librarians. For February, Web use statistics report the following:

91 page views, 85 user sessions, and 31 unique users. For March, the following was reported: 213 page views, 190 user sessions, and 34 unique users. Assessment of the guide is still in the early stages but the feedback so far indicates that the guide is a useful tool for accessing difficult-to-find information.

The Goddard Library has an intelligent, but very "junior," reference staff. The guide has aided them in their own personal exploration of the Goddard information landscape and has made it possible for them to contribute to it as well. The benefits of having Early's knowledge available in a written form can be summarized as follows:

1. *The knowledge can be more readily shared among a group of people rather than just benefiting a single individual.*
   When Early helps one of the staff with a question, only that staff member gains knowledge from the interaction. By compiling the knowledge in written form, all of the library staff will benefit because the resources necessary to answer the question will be at their fingertips. Not only will library staff benefit, but also the project librarians situated in many different locations on the Goddard campus. Where before only one person benefited, now dozens will benefit from Early's knowledge.
2. *Questions can be addressed in a timely manner.*
   Early is not always available when a patron calls for help. By taking a proactive approach, the staff are able to answer the patron's question quickly, even when Early is out of the office. Also, rather than spending hours making phone calls to discover where some archived information is stored (e.g., a print report from 20 years ago that hasn't been indexed anywhere), the staff can find the correct repository immediately.
3. *The* Goddard How To Get It Guide *will be the pooled knowledge of many brains, not just one.*
   The guide is not intended to represent just the experience of one reference librarian but to represent the experience of all the reference librarians as they learn and grow in their jobs and as information sources change or expand. Any one of the librarians can add an entry to the guide.

> Different people have their own specialized knowledge to which everyone will now have access, and, thus, everyone will benefit.

CONCLUSION

The library now has a dynamic, ever-growing knowledge base, which will improve the reference services of the library and project libraries by increasing awareness of the many resources available to us, and will also provide a training tool for new staff members. The ability to record and share knowledge with colleagues, regardless of location, is of vital importance in a desktop work environment. Human and information resources are geographically distributed and connected through technology. The guide serves as a tool for collaboration as well as a repository of collective reference knowledge.

The e-mail reference transactions precariously stored in the outbox of Early's e-mail account turned out to be an irreplaceable representation of his knowledge and his work. The importance of the tangible artifacts created from virtual reference transactions will no doubt continue to be an important topic of study. The lesson learned for the Goddard Library, and perhaps for other libraries, is to take advantage of technology to preserve the valuable knowledge contained within virtual transactions.

The *Goddard How to Get it Guide* took several months and the work of six people to create. While the effort has truly been that of a team and will continue to be so, the guide could not exist without the knowledge and experience of Charles Early. The guide will be renamed in recognition of his work but also to better capture the functionality of the guide. Henceforth, the guide shall be known as the Cooperative Holistic Approach to Retrieving Literature in Engineering and Science, CHARLES.

NOTE

1. The database was designed by staff member Lee Goldblatt. The guide is built on an AMP open source platform. AMP is an acronym for the combined use of an Apache Web server, MYSQL relational database, and PHP scripting language. This platform was chosen as it is rapidly becoming the de facto standard for serving dynamic Web pages to users. The advantages

of AMP include a global-scale user base, open source licensing, stable and scalable structure, and easy migration. The global user base has created copious documentation, and this vast community supports fast development, quick resolution of technological bugs, and no re-inventing of the wheel. The scalability of the platform is assured as it was designed from the ground up to handle heavy Internet work loads. Easy migration is made possible as all popular operating systems are supported. Therefore, an organization could change from a Linux to a Solaris host and keep the same AMP implementation. For more information on the AMP platform, www.onlamp.com/.

REFERENCES

Doezema, June C., and Barbara J. Fox. 1998. *How to Get It: A Guide to Defense-Related Information Resources*. 6th ed. Fort Belvoir, VA: Defense Technical Information Center.

Gandhi, Smiti. 2004. "Knowledge Management and Reference Services." *The Journal of Academic Librarianship* 30, no. 5 (September): 368–381.

Gross, Madeline M. 2002. "Capturing Undocumented Worker-Job-Knowledge at Electric Utilities: The EPRI Strategic Project." New Century, New Trends. Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants (Cat. No. CH37335). Scottsdale, Arizona: IEEE: 6–24.

Hirsh, Sandra, and Jamie Dinkelacker. 2004. "Seeking Information in Order to Produce Information: An Empirical Study at Hewlett Packard Labs." *Journal of the American Society for Information Science and Technology* 55, no. 9: 807–817.

Hoffman, Robert R., and Lewis F. Hanes. 2003. "The Boiled Frog Problem." *IEEE Intelligent Systems* (July/August): 68–71.

Jantz, Ron. 2001. "Knowledge Management in Academic Libraries: Special Tools and Processes to Support Information Professionals." *Reference Services Review* 29, no. 1: 33–39.

Katz, William A. 2002. *Introduction to Reference Work*. 8th ed. 2 vols. Boston: McGraw-Hill.

Liebowitz, Jay. 2002. "Knowledge Management in a Large Government Organization." *Information and Knowledge Sharing: Proceedings of the IASTED International Conference*, St. Thomas Virgin Islands: 51–53.

"OCLC Acquires 24/7 Reference." (2004). *Library Journal* 129, no. 15 (September): 22.

Perez, Ernest. 1999. "Knowledge Management in the Library—Not." *Database* 22, no. 2 (April/May): 75–78.

Stover, Mark. 2004. "Making Tacit Knowledge Explicit: The Ready Reference Database as Codified Knowledge." *Reference Services Review* 32, no. 2: 164–173.