

The Internet Library

Case Studies of Library Internet Management and Use

Edited by Julie Still

Mecklermedia
Westport • London

PLANNING BASICS FOR A LIBRARY ftp SITE: THE "INCREASE AND DIFFUSION OF KNOWLEDGE" AT SMITHSONIAN INSTITUTION LIBRARIES



MARTIN R. KALFATOVIC
Information Access Coordinator
Smithsonian Institution Libraries
libem071@sivm.si.edu

The quick and easy diffusion of information has long been a goal in the library profession. The rapid development of the Internet has made unprecedented amounts of material available to an increasingly large population. Libraries have taken advantage of this new technology by providing a wide array of services and products to their own patrons as well as to an international clientele.

Telnet access, via the Internet, to library catalogs and databases has received the most attention in the library world. File Transfer Protocol (*ftp*), though in many ways less glamorous, is a quick, efficient, and fairly simple way to make relatively static library information files available to the Internet-user population.

The technical planning for the creation of an *ftp* site is a fairly simple procedure; the intellectual processes of selection and authority of resources, however, can be much more complex.

SMITHSONIAN INSTITUTION LIBRARIES

The Smithsonian Institution was founded in 1846 under the terms of Englishman James Smithson's will. Smithson left his estate to the people of the United States for "the establishment, in Washington, of an institution for the increase and diffusion of knowledge." Early debate centered on the course Smithson's Institution should take. Some felt that a public museum would best serve the terms of Smithson's will; some, a scientific research center; yet others felt that it should serve as the basis for the creation of a large research library.

Today, nearly 150 years since its founding, the Smithsonian Institution reflects all these early concepts of what it should be. Smithsonian Institution Libraries (SIL), filling the role of the planned research library, is composed of eighteen branch libraries located in the museum and research facilities on and around the Mall in Washington, as well as at Smithsonian sites in the greater Washington area; New York City; Cambridge, Massachusetts; and the Republic of Panama. Supporting a wide range of research activities—from art history to zoology—as well as public programs, education, and a large publishing division, SIL has been described by Assistant Director of SIL Nancy Gwinn as being “afoot in three camps”: those of the academic, special, and museum library (Gwinn 1989).

Serving varied patron groups, from the museum-going public to postdoctoral researchers, spread over a wide geographic area, SIL has looked toward expanding access to its resources through various electronic means. One of those currently being explored is the use of an Internet anonymous ftp site.

THE INTERNET FILE TRANSFER PROTOCOL

The ftp has a long history with the first discussions of the concept originating in 1971 for work being done at M.I.T. (Dillon et al., 1993). Frequently discussed in the Request for Comment (RFC) literature over the next few years, the first “official” ftp document, RFC 454, was published in February 1973. Work continued on the ftp until June 1980, when RFC 765 was issued, defining the ftp specifications for TCP. Standards have continued to be refined since 1980, and RFC 959 (October 1985) is currently the official document for ftp specifications.

The ftp model, as defined by RFC 959 and by now familiar to millions of Internet users, may be described as follows: A user-ftp establishes a TCP/IP connection to a server-ftp. The user and the server can be any combination of platforms—Unix, DOS, Apple, and so on. The server-ftp is allowed to request from the user-ftp a user name, password, and account (Note: in the case of an anonymous ftp, the user name is “anonymous” and the password is the users’ electronic id; accounts are not generally required). Once the server-ftp has been satisfied by the user-ftp responses, the user is allowed to move around and within the host machines file structure and to retrieve and store files (as user privileges allow).

Because of the wide variety of computing platforms and the complexities that would arise with the necessity of site-specific

procedures for initiating file transfer between these platforms, the creation of the ftp protocol has established a baseline for this basic Internet function.

Ftp Statistics

In 1992, OCLC conducted a study entitled “Assessing Information on the Internet: Toward Providing Library Services for Computer-Mediated Communication.” The purpose of this study was to investigate “the nature of electronic textual information accessible via the Internet” (Dillon et al. 1993) and to determine the effectiveness of “cataloging or otherwise describing” (Dillon et al. 1993) textual resources on the Internet. In this general discussion of Internet resources, a number of interesting statistics on ftp use were revealed:

- * 47 percent of network traffic in bytes were for ftp date of information: November 1992 (Dillon et al. 1993)
- * 1,044 ftp sites with 3.1 million files existed date of information: Summer 1992 (Dillon et al. 1993)

In the results of statistical analyses and classification of these millions of ftp-able files, however, only a small percentage are the text and images files that a library site would likely produce. In random samples of twenty ftp sites, the OCLC report noted that only 7.3 to 7.4 percent were text; 1.9 to 2.4 percent were images, (Dillon et al. 1993). Even though the proportion of text and image files are small, a quick calculation will reveal that between 285,000 and 303,000 files fall into this category. The vast majority of files, over 50 percent, were system and source code files (Dillon et al. 1993).

THE SMITHSONIAN INSTITUTION LIBRARIES ftp PROJECT: GENESIS

The Museum Reference Center (MRC) branch of SIL serves an internal population including the Office of Museum Programs and acts as an important repository of material and information on museums and their collections. Among the material in the collection are unique bibliographies on various aspects of museology prepared under the direction of the branch librarian Sylvia Churgin.

In past years, photocopies of these bibliographies were mailed to researchers interested in them. After a posting on the Museum-l listserv in early 1993 that discussed the availability of these MRC bibliographies, Churgin was inundated with requests for copies. As

the bibliographies were in a machine-readable format and requests were coming from as far away as Australia, the librarian began to upload and send the bibliographies directly to the requestors via the Internet.

As demand continued, consultation with the Smithsonian's Office of Information Resource Management (OIRM),¹ indicated that the creation of an anonymous ftp site would cut down on the staff time needed to send (even electronically) the bibliographies and would also provide the mechanism for the diffusion of other SIL-produced information.

A proposal for creating an ftp site was sent to SIL's Director's Council. Preliminary approval was granted by Director's Council, and SIL's Public Program Committee, was deemed the appropriate group to begin discussing the implementation of an ftp site for SIL.

The Smithsonian Institution Libraries ftp Project: Concerns

Discussion within the committee focused on the primary issue of quality control (including issues of the continuing currency of information) of SIL information made available at the site. To help answer this and other questions, a literature search was undertaken to find out more about library implementation of ftp.

Perhaps the most interesting result of this search was that the library community has paid little attention to either of these issues. Gord Nickerson, writing in October 1991 (Nickerson 1991), discussed ftp, but only as a resource to be tapped by the library community, not one to be created. In a brief, but insightful article, Erik Jul, drawing on research done for an OCLC study (Dillon et al. 1993), again discussed ftp sites as repositories of information useful to the library community. Jul goes as far as to say that ftp represents the "largest electronic publishing system in the world" (Jul 1992).

Jul is not, however, uncritical of the varying quality of textual and image material available:

The degree of preparation evident in these ftp publications varies as widely as content and format, and if our worldwide ftp electronic publication system has one weakness, this is it.

Many documents reflect great care in preparation: the contents are complete, well written, free of errors, and formatted for ease of use. This represents the high end of the scale; other publications are more "casual" and prone to all the problems of amateur writers, editors, and publishers This is the ultimate personal publishing system, and all the foibles of personal publishing are to be found (Jul 1992)

Even with these reservations, he still sees ftp as "the primary method of electronic publication" (Jul 1992).

The Chronicle of Higher Education, responding to the increased importance of network activities, especially the Internet, ran a number of articles on the Net in 1993. Describing the tools available for Internet access, David L. Wilson states the "real power of the system [Internet] lies in the ability to remotely operate other computers . . . [and to] grab a copy of a file and move it to the user's home machine" (Wilson 1993a). Following up on this article a few months later, Wilson outlined some of the administrative fears that crop up when discussing the dissemination of information on the Internet. The most basic concern was, understandably, cost. Other worries included the fear that "valuable information available through the Internet is stored in ethereal form that may not be available in the next century or next week" (Wilson 1993b). The related issue of standardization and currency was also voiced: "people who want the most current version of a file . . . may accidentally get one that is years old and has never been updated" (Wilson 1993b).

Since no information addressing the concerns of SIL on ftp sites was located, a question was posted on the Public Access Catalog Systems listserv (PACS-L), asking for comments and information on policies and procedures that would address the issue of quality control. Responses were received from the Australian Nature Conservation Agency (ANCA) Library (Canberra) and the Lewis and Clark College Library (Portland, Oregon).² Comments from these institutions helped to clarify the ground to be observed when mounting library material on an ftp site.

Among the comments were the ideas that the material should be the following: the intellectual property of the library and thus freely reproducible; consistent with the purpose of the library and its parent institution; kept up to date and accurate; and that a representative group of library staff evaluates material placed on the site.

The Library Paradigm for Anonymous ftp Sites

Paradigms exist in the library community for ftp sites. The Library of Congress (LC) opened its ftp site to the public in the fall of 1992. The project is described as follows:

This is a pilot project from the Special Projects Office of the Library of Congress (LC). The goal of the project is to test the feasibility of making information from the Library available over the Internet These documents are only a small selection of

materials from various departments within the Library and are not considered to be representative in any way. Materials included here are public domain, generally produced by the designated office within the Library (Library of Congress 1992).

Though disclaiming the collection as a "small selection," the range of material made available (as of October 1993) is impressive. Information available includes exhibition catalogs (with image files) from such popular exhibits as "1492: An Ongoing Voyage," "Scrolls from the Dead Sea: The Ancient Library of Qumran and Modern Scholarship," "Revelations from the Russian Archives," and "Rome Reborn: The Vatican Library and Renaissance Culture." Documents related to special LC activities including FEDLINK (the Federal Library and Information Network) and FLICC (the Federal Library and Information Center Committee) were included. More common material such as brochures to its many readings rooms, subject bibliographies, and guides to its online catalog (LOCIS) were also made available.

In another recent example, the University Library of the University of Michigan (Ann Arbor) in association with the university's School of Information and Library Studies (SILS) established an ftp site as part of its Internet Resource Discovery project (IRD). Extending the bounds of supplying only local material by acting as a repository for guides to Internet resources on a wide variety of topics, the project aims to provide

. . . the University Library with these guides . . . and thus help extend the Library's collection development effort to electronic network-based media . . . to benefit members of the Internet community who may wish to consult these guides for personal use or for assistance in building their own subject-oriented servers (University of Michigan 1993).

With these and a growing number of library examples of ftp sites being created on the Internet, SIL began the process of creating policy parameters for material for its own ftp site.

Conclusion

The explosive growth of information on the Internet is reflective of both increased accessibility to the network by greater numbers of people and the urge by these users to use and create information.

One scholar termed the library a "gigantic mincing-machine, into which the labours of the past are flung, to be turned out again in a slightly altered form as the literature of the present" (Clark 1894). As libraries and librarians are able to create the raw materials for future information, they must be careful to integrate their traditional role of repository and access point with that of creators. In the past, any publishing done was usually in the form of material for in-house use that saw little outside circulation. With the advent of electronic publishing in the form of ftp, which allows for instantaneous and widespread dissemination of library publications, libraries must add the role of publisher to their traditional role of information repositories.

Internet tools such as Gopher, Mosaic, and WAIS will make information available on the Internet retrievable by a less-sophisticated clientele. Ftp will remain, however, the basis for the retrieval of static information by these tools. For the foreseeable future, a portion of the Internet community will lack these tools (just as now some lack ftp access); for this segment of the population, ftp will be a valuable information-access device.

Additionally, the relative ease of ftp for experienced Internet users makes the retrieval of even large files a fast and easy process. When the location, directory, and filename are known, even newcomers to the Internet can quickly bring information from around the world to their desktop PC.

Project Update

Policy guidelines for SIL material to be made available via ftp are currently being drafted. In the interim, selected material is being placed in the Smithsonian directory at simsc.si.edu.

Notes

My thanks to Barbara J. Smith, director; Bonita D. Perry, assistant director for research services; Nancy L. Matthews, publications officer (all Smithsonian Institution Libraries); and Julie Cochran PROFS/OV Administrator (Office of Information Resource Management).

1. OIRM fulfills many of the functions that a university's computer center would control.
2. My thanks to Roxanne Missingham, librarian in charge, Australian Nature Conservation Agency, and Beverly B. Stafford of the Lewis and Clark College Library, for sharing their views and comments on this issue.

Bibliography

- Clark, J. W. *Libraries in the Medieval and Renaissance Periods*. Cambridge: Macmillan and Bowes, 1894.
- Dillon, M., E. Jul, M. Burge, and C. Hickey. *Assessing Information on the Internet: Toward Providing Library Services for Computer-Mediated Communication*. Dublin, OH: OCLC Inc., 1993.
- Gwinn, N. "The Smithsonian Institution Libraries: Afoot in Three Camps." *College & Research Libraries* (March 1989): 206-14.
- Jul, E. "Ftp: Full-Text Publishing?" *Computers in Libraries* 12(5) (May 1992): 41-42.
- Library of Congress. WELCOME to the anonymous ftp server at the Library of Congress. (Document available via anonymous ftp. Site: ftp.loc.gov; Directory: pub/Library.of.Congress; Filename: README.) September 9, 1992.
- Nickerson, G. "Networked Resources: File Transfer Protocol." *Computers in Libraries* 11(11) (October 1991): 51-53.
- Postel, J. and J. Reynolds. File Transfer Protocol (ftp): Request for Comments 959. (Document available via anonymous ftp from many sites.) October 1985.
- University of Michigan, University Library and the School of Information and Library Studies. About the Internet Resource Discovery Project (UMich). (Document available on UMich Gopher.) August 1993.
- Wilson, D. L. "Array of New Tools Is Designed to Make It Easier to Find and Retrieve Information on the Internet." *The Chronicle of Higher Education* (May 26, 1993a): A17-19.
- Wilson, D. L. "Electronic Riches Are Free on the Internet, but Some Worry About the Consequences." *The Chronicle of Higher Education* (July 28, 1993b): A18, A20-21.