Supplementing the Library Collection with Digital Content from Engineering Departments

Karen Clay
Engineering Library, Stanford University

There are over 65 Centers and Labs associated with the Stanford School of Engineering. Information from these Labs is often being made available online, either in addition to or instead of print. Much of this online material is widely scattered, difficult to find, mounted on personal websites, and unlikely to be available over the long term. Ideally, the research output of the School of Engineering Labs and Centers, whether in print or electronic format, should be preserved for the long term and should be easily accessible to the Stanford community. As a first step towards this goal, the Engineering Library has begun a project to identify, characterize, and organize these materials. We want to learn what is out there, where it is being stored, and how much of it we already have in the Library.

Documents were identified by systematically searching Stanford School of Engineering web space, and by contacting each of the School of Engineering Labs and Centers individually and asking them if they could send us an inventory of their research output. For every item found with substantive informational content, descriptive information was gathered and input into a bibliographic database. The database was created using Refworks, an internet based bibliographic management software package.

The information input for each document includes bibliographic information, abstract and descriptors if available, the affiliated School of Engineering Center of Lab, the document type, document format, document size, and URL if available. We are also noting whether or not a record for the item exists in the Library OPAC, and if so, the call number. Figure 1 below shows a typical record in the Refworks database.

Copyright © 2004, American Society for Engineering Education"

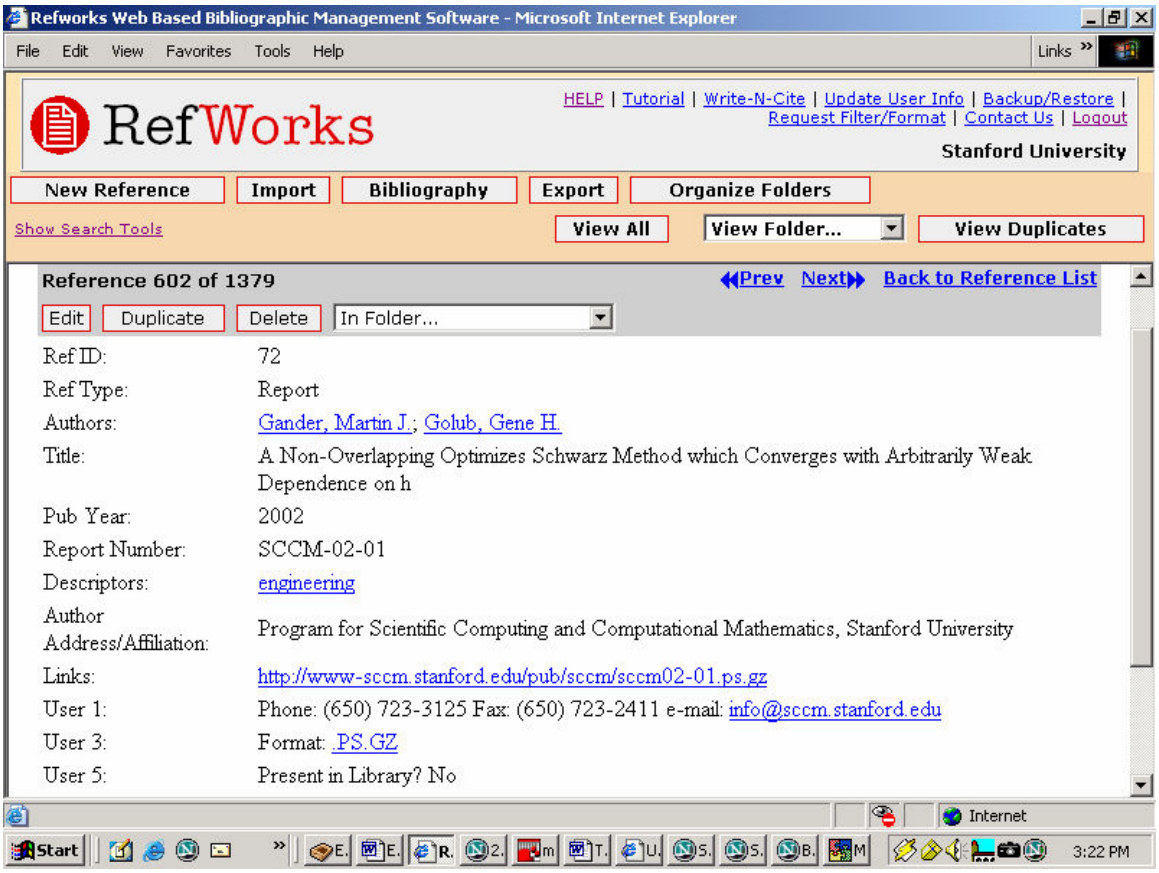

Figure 1. Example of a record in the Refworks database

We identified 66 research centers affiliated with the Stanford School of Engineering, and data has been collected for 23 of these. The Refworks database contains 2052 records, each representing a document produced by a Lab or Center. Table 1 lists the Centers which have some content in the database and shows how many records were input for each Center. Many have very few records, while for others, notably the Knowledge Systems Laboratory, the Center for Turbulence Research and the Center for Integrated Facility Engineering, we created hundreds of records.

Copyright © 2004, American Society for Engineering Education"

Name of Lab or Center	Number of records in Refworks
Aero Fluids Mechanics Laboratory (AFML)	2
Aerospace Robotics Laboratory (ARL)	11
Alliance for Innovative Manufacturing at Stanford (AIM)	3
Biomotion Research Group	102
Cardiovascular Biomechanics Research Laboratory	19
Center for Design Research (CDR)	63
Center for Integrated Facility Engineering (CIFE)	250
Center for Integrated Systems (CIS)	6
Center for Integrated Turbulence Simulations (CITS)	5
Center for Polymer Interfaces and Macromolecular Assemblies (CPIMA)	5
Center for Turbulence Research (CTR)	369
Engineering Risk Research Group (ERRG)	20
Environmental Fluid Mechanics Laboratory (EFML)	167
Global Positioning System Laboratory (GPS)	75
High Temperature Gasdynamics Laboratory (HTGL)	52
John A. Blume Earthquake Engineering Center	138
Knowledge Systems Laboratory (KSL)	459
Rapid Prototyping Laboratory (RPL)	96
Scientific Computing & Computational Mathematics (SCCM)	163
Stanford Technology Ventures Program (STVP)	12
Structures and Composites Laboratory	13
Systems Optimization Laboratory (SOL)	9
Western Regional Hazardous Substance Research Center	13

Table 1. Number of records input for each of the Engineering centers in the project

One use for this data was to help us discover what types of digital materials are being produced by the research centers. As making materials available on the web has become simpler for the average individual or institution, it has also become more common. It has been postulated that the web is having a profound effect on the availability of information which does not necessarily follow a standard, well-established route to publication[3,5]. While this type of information (sometimes referred to as grey literature or ephemeral literature) can often pose difficult collections challenges, its ubiquitous presence on web pages is helping to propel it to greater prominence in the eyes of both librarians and users.

Copyright © 2004, American Society for Engineering Education"

It was therefore suspected that this project would reveal a large amount and variety of grey literature, particularly digital, web-based grey literature. These types of materials have high information value and would be a valuable supplement to the library collection, but they are often more difficult to acquire in physical format, as they are not restricted to following the usual channels by which published materials are produced.

Once the records were examined we found that most documents fell into clearly defined categories such as dissertations, technical reports, conference or meetings papers, book chapters and journal articles. Other material types are videos or dvds, presentation slides, and newsletters. Web pages which did not fit into any of the existing classifications were simply grouped together under the category "web pages". There were very few of these in fact (only 4) and they all consisted of listings – listings of software programs, an event archives, a workshop archives, and technical highlights.

In addition to analyzing the material types, we also determined which materials had been made available online, and how many were available in the Library. All the records in the Refworks database were assigned a URL if an online version could be identified, and availability in the Library was ascertained by searching for each document in the library catalog. Table 2 shows a breakdown by document type for all the materials in the Refworks database, for the materials that are available online, and for the materials that are available in the library.

Document Type	# documents in total	# documents available online	# documents available in the Library
All Documents	2052	1379 (67%)	962 (47%)
Book Chapters	8	3 (38%)	5 (63%)
Conference Papers	654	524 (80%)	95 (15%)
Dissertations	93	30 (32%)	87 (94%)
Journal Articles	188	26 (14%)	176 (94%)
Newsletters	4	4 (100%)	0
Presentation slides	13	9 (69%)	0
Technical Reports	1051	761 (72%)	600 (57%)
Videos or DVDs	37	14 (38%)	0
Web pages	4	4 (100%)	0

Table 2. Breakdown by document type for all materials, online materials, and materials in the Library

"Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition Copyright © 2004, American Society for Engineering Education"

It is apparent from our results that a large proportion of the material is grey literature. That is, literature which is not published by mainstream publishers, and which is often characterized by uncertain availability, poor bibliographic control, and non-professional appearance[1]. Many librarians[2,4,6] have noted that while grey literature is critical for engineering research, it is not always easy to identify or to obtain. For the purposes of this project the web pages, videos, technical reports, presentation slides, newsletters, dissertations and many of the conference papers (those not published in easily purchased proceedings volumes) are all considered grey literature. These comprise 1752 records or approximately 85 percent of the records in the database. Given the extra effort usually needed in order to ensure that grey literature is incorporated into the library collection, it would not be too surprising to find that many of materials are not in the Libraries' collection.

In terms of numbers of materials, technical reports predominate, followed by conference papers. Although both of these materials types are considered grey literature, they are established grey literature and institutions and methods have evolved to keep track of them. The more ephemeral grey literature (the web pages, newsletters, videos and presentation slides) would be much more difficult for a library to systematically collect and make available. Because the web has made the production and presentation of such documents much simpler, it was expected that this type of literature would predominate. In fact, although there is a wide variety of materials that would be considered ephemeral, in terms of number they only comprise 58 documents, less than 3 percent of all documents identified.

Another notable finding is that neither the web nor the Libraries are proving to be comprehensive sources for the identified documents. Only 67% of the documents in the Refworks database were available online, and after 5 months 13 URLs had either changed or disappeared. The document types where the majority of the materials are available online are the conference papers, technical reports and presentation slides, web pages and newsletters. It could be that this is because these documents are easy to mount online from a technical standpoint as they are often produced digitally, and also they pose fewer copyright restrictions than the more formally published journal articles and book chapters.

The Libraries at Stanford University held only 47% of the materials identified in the Refworks database. Although this result was somewhat expected, as grey literature can be difficult to obtain, the extent of the gaps in library holdings is significant. Library holdings for the different document types were therefore examined in more detail.

Copyright © 2004, American Society for Engineering Education"

Journal Articles and Book Chapters:

Library holdings for these document types are relatively comprehensive - most of journal articles and book chapters in the Reworks database are from publications held in the Libraries. There are only 12 journal articles and 3 book chapters from journals or books not held in the Library. Although this is not a high number, it is worth looking at the records in more detail to see if there are publications that show up more than once – that is, journals or books which contain a more than one article produced by School of Engineering Labs or Centers. Once identified, these should be added to the library collection. In fact, the 12 journal articles and 3 book chapters are all from different publications. One significant finding is that 8 of the 12 journal titles and 2 of the 3 books are in the biomedical field. Bio-engineering research is blossoming at Stanford with the new BioX program and the recent formation of the BioEngineering department, and this result is an early indication that more information resources will be needed in that area in the very near future.

Conference proceedings:

When the 558 conference papers that are not in the library collection were examined in more detail, it was found that 369 of the records were part of an annual summer conference held by the Center for Turbulence Research. The conference proceedings are not high-profile and unlikely to be available through regular channels. The straightforward expedient of obtaining a copy of the proceedings directly from the Center and adding it to the Library collection would provide access for over half the conference proceedings identified as not in the library collection. Nine other conference proceedings were identified containing 5 or more records from the database, and these should be assessed on an individual basis for possible addition to the Library collection.

Dissertations and Technical Reports:

The Libraries have procedures in place intended to ensure that we receive a copy of every technical report and every dissertation produced at Stanford. These procedures work well for the dissertations. Of the 6 dissertations not available in the Libraries, 3 are non-Stanford dissertations, and the other 3 are Masters theses. The same cannot be said of the technical reports. The Libraries are missing technical reports produced by 14 of the different labs and centers, with significant gaps in the Libraries holdings for several of these. This lack could be attributed to a breakdown in communication with the labs and centers, a failure of the established procedures to continue over years of staff renewal. It may also reflect an attitude that with web delivery being so simple, the library is no longer needed as a storage place. The best way to tackle these gaps would be to initiate and maintain communications with the Labs and Centers for which we are missing reports.

Copyright © 2004, American Society for Engineering Education"

Conclusion

The technical and logistical barriers to making information available over the web have all but evaporated in recent years. This being the case, we expected the project to yield a significant number and variety of ephemeral, web-only type documents. However, for the present at least, over 97% of the materials identified for this project could be considered "traditional" publications formats. These formats can be difficult to obtain, but they do not pose any new collections challenges. It is likely that the non-traditional, often web-only, document types were less numerous than expected because the whole process of scholarly communication so important to academic success is still very closely tied to the more traditional publications types.

Most of the document types identified during this project could be considered grey literature, with the standard formats of conference papers and technical reports predominating. The relatively few web-only items were usually web listings, laboratory newsletters, or presentation slides. This finding has implications for library collections in that the more ephemeral forms of grey literature can often pose the most difficult collections challenges. The conference papers and technical reports (which in this case predominate) are more straightforward to obtain as they are usually available through well-established channels.

Information produced at Stanford should be available to the Stanford community, but locating the full text of many of the documents listed in the Refworks database is not a straightforward task. The web was far from a comprehensive source for these documents, as only two thirds of the documents listed had URLs, and only 5 months after the project was done, some of these web links have changed or disappeared entirely.

Stanford University Libraries are unfortunately an even poorer source, with less than 50% of the Refworks listings having a record in the library catalog. More detailed comparisons between the Libraries holdings and the Refworks database indicate that the Libraries do an excellent job of collecting the standard literature such as journal articles and a very thorough job of collecting Stanford produced dissertations. Despite the mechanisms in place to collect technical reports, the Libraries are lacking over 40% of the technical reports in the Refworks database, a much higher than expected proportion. The other document type very poorly represented in the Libraries' collection is conference proceedings.

This exercise proved to be a very useful check on the Libraries' engineering collection. Arguably our most significant finding is that most of the documents identified and entered into the Refworks database are not physically held in the Engineering Library and have no record in the online catalog. This implies that there are hundreds of currently available, relevant online documents literally at our back door, which should be incorporated into the Library collection. Library involvement in this area is badly needed, as the web is not proving to be a reliable long-term source for local information, and only proactive efforts on the part of librarians will ensure that valuable information is not lost.

Copyright © 2004, American Society for Engineering Education"

Bibliography

1. Auger, C.P. 1998. *Information Sources in Grey Literature*, 4th ed. New Providence, N.J.: Bowker-Saur.

2. Chan, Winnie S. and Deborah Rhue. 2001. Providing a Digital Portal to a Print Collection: A Case Study for an Engineering Documents Collection. *Science and Technology Libraries*, 21(1/2): 29-42.

3. Gelfand, Julia M. 1998. Teaching and exposing grey literature: What the Information Profession Needs to Know – Examples from the Sciences. *Collection Building*, 17(4): 159-166.

4. Gelfand, Julia M. 1999. Grey Literature Poses New Challenges for Research Libraries. *Collection Management*, 24(1/2): 137-147.

5. Primack, Alice L.1989. Identifying Academic Physical Sciences Research in Progress: Current Sources and a Proposed New Database. *Collection Building*, 10(1/2): 22-28.

6. Thompson, Larry A. 2001. Grey Literature in Engineering. *Science and Technology Libraries*, 19(3/4): 57-63.

Biography

KAREN CLAY. Karen is the Head of the Engineering Library at Stanford University.

Copyright © 2004, American Society for Engineering Education"