

Summer 1999

Integrating Digital Resources into a Traditional University Research Library

Fiona C. Coutinho

Caroline M. Eastman

University of South Carolina - Columbia, eastman@cec.sc.edu

Christopher B. Hare

University of South Carolina - Columbia, chris@mailbox.sc.edu

Robert F. Skinder

University of South Carolina - Columbia

Follow this and additional works at: https://scholarcommons.sc.edu/lib_facpub



Part of the [Library and Information Science Commons](#)

Publication Info

Published in *Issues in Science and Technology Librarianship*, Volume 23, Summer 1999.

<http://www.istl.org/99-summer/article3.html>

© 1999 by Association of College and Research Libraries

This Article is brought to you by the University Libraries at Scholar Commons. It has been accepted for inclusion in Faculty and Staff Publications by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

Integrating Digital Resources into a Traditional University Research Library

Fiona C. Coutinho

Department of Computer Science
University of South Carolina
Columbia, SC 29208

Caroline M. Eastman

Department of Computer Science
University of South Carolina
Columbia, SC 29208

Christopher B. Hare

Thomas Cooper Library
University of South Carolina
Columbia, SC 29208

Robert F. Skinder

Thomas Cooper Library
University of South Carolina
Columbia, SC 29208

Abstract

We describe the ongoing *Electronic Library Project* at the University of South Carolina. The goal of this project is the integration of digital resources within a traditional university research library. The first step was the development of an *Electronic Science Library (ESL)*, followed by an *Electronic Academic Library (EAL)* which includes non-science subjects. We discuss the structure of these libraries and comment on our experiences with their implementation and use. The prototype implementation used static web pages, a technology which we knew would not scale up well. This implementation is being replaced by a database system using SQL Server and Active Server Pages. Future plans are briefly discussed.

Introduction

Every year Internet-based resources become increasingly important to the academic community, but they do not necessarily become easier to use or find. This paper

describes the *Electronic Library Project* being developed at the University of South Carolina. The objective is to harvest thousands of useful academic resources that are available on the Internet and make them available to the academic community in the form of an academic digital library. This is an ongoing project of which various components have been completed. We first discuss the *Electronic Science Library (ESL)* and the subsequent *Electronic Academic Library (EAL)*. We then describe our supporting database and retrieval system. Planned future work is briefly summarized.

Fundamental to our plan is a system that targets discrete items and places them within the structure of an electronic library. The items are placed on clearly marked shelves and assigned call numbers, thus providing structured access to the available but disorganized academic resources on the Internet. In this manner we make them available to faculty and students for teaching and research. To achieve that goal, we have laid out a multi-part plan, parts of which have been completed while others have yet to begin. The *ESL* was the first step and has taught us a great deal. One of the key lessons that we learned was that the human intellect is better used developing tools to produce electronic libraries than searching for the resources to be included. This is illustrated by the development of our database shell and retrieval engine.

The *ESL* was constructed and we learned from it. When we proposed the non-science *Electronic Academic Library (EAL)* we not only expanded our scope but went beyond the library to develop computerized solutions. As time passed, new applications for related projects have emerged, such as a project involving Internet-based biological databases. New computer-based tools have also been developed or proposed and will be discussed. In short, our original work has now become a test bed for a wide range of topics.

The overall goal has been to design a system that will produce an easy and efficient way for the academic community to use the resources of the Internet. The librarians are concerned with organizing the materials. Our Computer Science allies hope to utilize state of the art technology to reduce the amount of tedious labor inherent in such an endeavor. To date, these innovations include an ASP (Active Server Pages) based database coupled with a sophisticated retrieval engine. Additional tools will include a customizable search robot for discovering resources as well as computerized "Library Assistant" programs to manage housekeeping routines. These will be implemented as opportunities appear.

Generally speaking our approach has been for the librarians to build small segments of the library and observe both the researchers and the users. As problems or opportunities are discovered we confer with faculty of the Computer Science Department who have often been able to provide both answers and student programmers. Projects that have involved the Computer Science students to date are

the prototype database shell with search engine. They are beginning work with the spiders as we begin our biological database project.

Electronic Science Library

Objectives

The prototype *Electronic Science Library (ESL)* was developed to determine if it was feasible to locate, evaluate and make available to the classroom and laboratory the numerous educational resources available on the Internet. Limiting our initial efforts to science seemed to be a manageable goal, and the initial investigators were science librarians. It was, however, always assumed that the work would be expanded given any positive results. Secondary objectives included:

- Determine the best way to locate these resources.
- Evaluate the quantity and quality of these resources.
- Devise a system for housing these resources that would be conducive to our arranging them and our patrons locating them.

The working model of the *ESL* can be viewed at <http://www.sc.edu/library/science/elibind.html>

Approach

The *ESL*, like most libraries, consists of two distinct parts. First is the infrastructure comprising the physical plant, the computerized system(s), the materials and the librarians. The other half of the equation consists of the users and those parts of the system that allow them to use the library. The approach therefore is divided.

1. Management and Development

Electronic or digital libraries include a wide range of concepts such as digitizing existing copies of documents or collecting materials that are already digitized such as maps or statistical data. Some digital collections concentrate on cultural or geographical highlights of the parent organization. Our particular strategy was to locate discrete, catalogable Internet resources such as particular books, journals or online courses. This concept is opposed to the "list of lists" approach or the use of search engines. These methods might be suitable for use by some professionals but are not sufficient by themselves for the undergraduate community.

Our resources were collected and arranged by two information specialists working from 5 to 10 hours a week for one year. Resources were located using the very tools that the *ESL* was designed to replace. The searchers used search engines, virtual libraries, mailing list messages, and Internet surfing. The information specialists used cataloging information when it was available. If not, they approximated call numbers and assigned appropriate locations. The work that they did was extremely labor intensive and turned out to be one of the major problems in exploitation of digital resources.

At the completion of our prototype, the format of our pages or shelves was still very much the same as when we began our work, which suggests that we guessed well. We have identified and entered into the database approximately 2,500 resources.

2. Patron Usage

To locate materials, the user chooses one of the sciences on the opening or {**index page**}. The subjects listed are not all of the sciences taught at USC, Columbia, but represent the larger departments and, perhaps more importantly, those that have a reasonable number of electronic resources available on the Internet.

They are then taken to the next page or area that is functionally always the same regardless of the subject chosen. Here are listed the separate {**categories**} of items. In the science library, these categories are the same for each subject. This allows the user to feel comfortable wherever they may be within the *ESL*. Please note that they vary considerably among the non-science *EAL* pages.

The individual categories are generally self-explanatory. The first two areas, *Department Home Pages* and *Faculty*, are included because the *ESL* is intended to be a part of the educational process at a specific institution. The *Online Courses* category reinforce this but also offers the students many opportunities to look at classes similar to, but not exactly the same as, those that they may be taking. We expect the online courses area to grow very rapidly.

Full-text Books are exactly that; when you select a particular title, the full text of the book appears. The National Academy Press has placed a large percentage of its books on line. There are also archival books whose copyrights have expired as well as several that are online due to the beneficence of the authors. There are also several full-text books that actually appear in other

sections such as Reference or may be found under another subject. There are far more full-text books available in the non-science areas. The exact opposite situation appears to exist with the online journals.

Online Journals are not as clear-cut an issue as the full-text books. As you know, this is the result of the many academic and societal publishers who seem unable to deal with the myriad issues associated with electronic publishing. As a result we have journals whose access ranges from full-text to abstracts or tables of contents. Since the number of full-text journals is growing so steadily, we will be giving serious consideration to eliminating the products that provide only a table of contents or abstract.

The *Reference* section currently consists of dictionaries, glossaries and encyclopedias. These are often divided by disciplines within a subject for ease of use.

The next line consists of *Search Tools*, *Databases*, and *Calculators and Tools*. The search tools are not the typical search engines associated with the Internet such as Yahoo and Excite but are Internet-based tools associated with a particular field such as Medline in the medical field or NASA Reports in the engineering sections. These provide an enormous wealth of information through, rather than within, the Internet.

Databases offer enormous potential in the classroom but they also present problems. One that is particularly obvious is differentiation between a search engine and a database. Although we have placed them in separate categories, it is increasingly apparent that the difference is relatively minor and they will soon be combined.

The third section of row 3 is called *Calculators and Tools* and consists of 3 primary areas. The first of these comprises a number of scientific calculators that have been developed using Java technology and that are available on the web at no cost. Applets, also made possible by Java, include virtual experiments and can be found primarily in the physics and astronomy sections. A typical applet displays one or more graphs or some other way to illustrate data points. An area to enter or change numerical values is also provided. Changing these values will immediately change the voltage, the trajectory, the orbit or any other value-dependent variable.

The final part of *Calculators and Tools* is a page designed to go beyond the discrete, catalogable item by pointing to important information regardless of where it is housed. Broad classes of discipline-specific information needs and

appropriate resources are identified. This is an experimental tool, designed by the Science Library, and is aimed at less sophisticated students who might not recognize that a particular tool or device was available on the Internet. This particular effort first appeared as the **Prototype Chemistry Page** but there are obviously many other applications where it can be used. The referenced page is for illustration only and shows only a few of the actual tools available.

The remaining three categories on row four are of minimal interest. Primarily they are pointers to areas that we have chosen not to develop fully but that are important enough to be noted.

Incidentally, work on the prototype *ESL* took approximately one year or 500 hours that included searching for resources and cataloging as well as routine housekeeping chores such as ensuring that URLs remain stable. Other tasks included developmental work in non-science areas and in the design and construction of a new database.

Results

The immediate results are a set of tools, useful for faculty, students and researchers that brings thousands of valuable Internet resources to them in a logical and easy-to-use format. Those are the visible results. We have also gained an understanding of what really is available on the Internet as well as the values and limitations associated with those resources, most notably the arduous work involved in a project of this nature.

The *Electronic Science Library* has often been used as the core around which Internet classes are built or it is cited as an important resource. These classes range from University 101 (an immigration course for new undergraduates) through graduate level seminars on information retrieval. We also introduce the *ESL* in all library information classes whether they are Internet-based or not. The design of the system in its present configuration is particularly well suited to the library classroom because we need only to pick the subject under discussion and from there, each category (books, journals, etc) is merely a click away.

The project has also allowed librarians to collaborate with teaching and research faculty for projects, classes and proposals. For this work to reach its full potential participation by the faculty is required and this is beginning to occur. Seminars, workshops and regular library newsletters will increase faculty and student awareness in the future. We are also making the *ESL* more visible by integrating important research tools such as the Web of Science, Lexis-Nexis and Current Contents with it.

On the negative side, this work has proven to be far too labor intensive. Searchers first must find a candidate object, then examine it. At a minimum they will then add it to a given page, making sure that the connection works. In many cases, depending on the category of material, call numbers, keyword, author and publisher's name and availability of a hard copy in the library were included. In some cases the same record might then be added to different pages. A good hour might possibly produce two or three completed items. Upkeep of the system has also proven to be a problem, particularly in the area of electronic journals where the publishers have continually developed new and trial pricing arrangements. We have had several occasions where one such arrangement causes a hundred or more journals to be added or to have their status changed.

Before the prototype was finished it was determined that the results warranted the continuation of our work but that modifications need to be made to our mode of operation, namely to facilitate the librarian's task by developing automated "Librarian Assistants." These, we hope, will increase productivity by assisting in all or part of some of the necessary tasks such as acquisition, search, retrieval, classification and indexing of resources. Software will also be used to simplify the users involvement by including search capabilities and arranging materials in a logical and user-friendly manner.

Database Development

Objective

The objective of this part of the project was to create a database, searchable on the web, to replace the numerous HTML links created in the initial phase of the ESL project. That database would include selected discrete, catalogable items, web and non-web items. Subject specialists can use the system to create and modify discipline-specific databases.

As previously stated, constructing the original web links turned out to be very labor intensive. Any one item may have been set up with numerous links. In addition, there were only predefined paths to subjects and no way to search by keywords or subject headings. Each item in the new database would be assigned limited subject headings to minimize the work of the selecting subject specialist. Subject headings could be set up, using an established controlled vocabulary, by the manager of the database or by the subject specialists themselves. At this time it appears that we will use the Dublin Core Metadata format modified to accept our nested subject headings.

Approach

An existing Microsoft NT Server, SQL Server and Internet Information Server installation was used. This avoided additional software and hardware costs and minimized the need for additional support personnel. In addition, only HTML, ASP and Visual Basic Script were used to keep the skill level needed for support minimal.

The project was broken down into three tasks:

- A. **Create a database with a table for each category.** All categories, whether they are online classes, books, journals or calculators, will be assigned the following Dublin Core Metadata fields: title, creator, subject (nested) and keywords, description, publisher, other contributors, date of publication, type of resource, format, URL, source, language, relation coverage, and rights holder. In our case we also intend to include the applicable discipline or disciplines.

Distilling and then building the database will remain our largest challenge. We are currently experimenting with manual procedures but have no doubt that our success relies on complete or partial automation of the process. Our future plans call for the development of such tools. Anyone who is involved in a similar project should look at the [DC-dot-Dublin Core Generator](#). This, at least, offers some help in developing the metadata. It works extremely well when the metadata have previously been developed and applied to a given item. In the near future, it is hoped that everyone who is developing an online resource will insure that each page has the core items installed or that the proper information is included in the design of the product so that meaningful metadata might be developed.

- B. **Create a web-based staff module with input and editing screens for each category.**

Upon reaching maturity, our system will automatically identify and locate those fields that constitute the database. In the meantime, input to editing screens will be done by humans with some subject expertise. The principle tasks will be to enter new records, create/delete disciplines, create/delete any of the core fields and edit existing records. This portion of the system has been developed and we will be entering data relative to our biology database project by the time that this paper is published.

- C. **Create a web-based patron module with search options.**

This portion has also been developed and awaits the input of data. It is designed to be integrated with other electronic projects and search tools. Its design is intended to meet the requirements of all searchers regardless of experience. In fact the system will perform best for the more inexperienced users. Search

options include selection of one or more categories, selection of a discipline, subject search, keyword search and call number search. One of the more interesting features is the nested subject search. A major subject heading will reveal all of the subject headings subordinate to it to three levels.

Results

The result of this work was the creation of a database shell. The shell allows a librarian to easily create a World Wide Web searchable database on any discipline. The collections would not be everything located on the web claiming to cover a discipline or subject heading, as traditional web search engines now produce. The collection model resembles the collection management policies of traditional librarians. Collections can be tailored to fit the needs of the curriculum taught at the University or to fill the needs of faculty and students based on their requests.

Searching is quite flexible. All hits that are on the web are presented as links. The searcher can choose one, a few or all the categories to search at the same time. Each subject search returns hits for all records for which that subject is the last heading in the record. In addition each subject search returns a list of all subject headings nested within it. Keyword searches query all fields in the selected category(s) and discipline. If a keyword happens to be a subject heading, the searcher will be given both keyword hits and subject headings nested within term. Call number searches can be used to browse related material. Wild card capability is built in.

Conclusion

We have attempted to provide a look at one university's attempt to integrate the new and ever increasing digital resources of the Internet with the traditional resources. This is not only a physical matter but also a psychological one. A considerable amount of inertia exists in both directions regarding Internet usage. Many people, including some librarians, are loath to consider an Internet resource a reliable one. Conversely many students spend hours looking for non-existent answers on the Internet when the actual answers are available on the library's shelves.

The physical portion of our project began with a fairly simple goal. We would evaluate resources available on the Internet. We would seek a sensible way to locate, arrange and use them and we would attempt to introduce them into the classroom and laboratory. Because we had limited resources, we started with the sciences and within one year (and <\$5,000) developed the *Electronic Science Library*. We learned important lessons from this effort. The resources were more plentiful and more useful than we expected. The work to locate them and assign order to them was also more

than we expected. Our conclusion was to proceed but to change the focus. Instead of using humans to search for and arrange the resources, we would use humans to design the tools for searching and to evaluate and teach the retrieved resources.

Exploratory work began on the Academic *EAL* before the *ESL* was near completion. It was important to have an idea of the quality and quantity of the non-science resources if we were to continue. In many ways the non-science resources differed from the scientific ones but there was no perceived decrease in value. We have begun construction of the *Electronic Academic Library (EAL)*. Graduate students in the College of Library and Information Science recently completed preliminary pages for English Literature, Library Science, Music and Philosophy. A collection for Statistics has also been placed in both the *ESL* and *EAL*. Work in non-science areas is as interesting as our initial endeavors with certain variations. Instead of databases and calculators we are finding materials such as interviews, artwork and sheet music. It appears, however, that the *EAL* will now pass to the control of another group within this library. The original designers will assist as consultants. At this point it seems likely that materials in the *EAL* will be entered into the University OPAC rather than into the database that we designed.

Non-science subjects currently available are:

- {English Literature}
- {Library Science}
- {Music}
- {Philosophy}
- {Statistics}

In closing, what began as an experiment now stands as a sturdy tool at a fraction of the cost of commercial databases. Librarians, faculty and students are becoming more familiar each day using and working with digital resources from around the world. After two years of development we have thousands of separate resources in an orderly format, as well as important computerized tools and techniques. We are in a good position to expand the scope of our system by a factor of at least five within the next year and continue with substantial growth for the foreseeable future. This work also serves as the basis for additional opportunities including those involving other departments, libraries and companies.

Future Projects

At the present time, Summer 1999, the *Electronic Science Library* is well established. A model for the academic library has been established and will be passed on to others. The database project is undergoing modification to adhere to the Dublin Core

Metadata concept and the search engine is ready for use. We are currently working on a project somewhat smaller in scope, the Biological Database Project. Two of the most important aspects of this project will be the automated spider that will find databases and our ability to immediately develop and impose metadata from the newly found object to our database of databases. Both of these projects will depend on automation, which will probably be accomplished in several steps. The value of attempting these automating projects in conjunction with the Biological Database Project is that everything is starting from ground zero. We will not be facing the prospect of developing and installing additional fields of information on a database that already holds several hundreds or thousands of items.

Acknowledgment. Grateful appreciation to BellSouth Instructional Innovation Grants, 1997 and 1999.