Proceedings of the 8th Annual
Federal Depository Library Conference

April 12 - 15, 1999

Digitizing Collections of Government Documents: Options, Processes, and Costs

Cathy Nelson Hartman, University of North Texas Libraries
Denton, TX

Introduction

Good afternoon. My name is Cathy Hartman, and I am the electronic resources coordinator for government information at the University of North Texas (UNT). My journey down the path to digitization began in 1997 at the ALA Midwinter Meeting in Washington, DC. Duncan Aldrich, then working as an expert consultant in GPO's electronic transition effort, mentioned in an update at GODORT's Federal Documents Task Force meeting that GPO needed depository library partners to assume responsibility for providing access to electronic documents. Depository libraries have provided access to documents in tangible media for more than a hundred years, so providing access to documents in electronic media seemed to be just another method of fulfilling our responsibilities as a depository library. I volunteered to become a partner.

In October 1997, the University of North Texas Libraries entered into an agreement with the U.S. Government Printing Office to provide permanent public access to the electronic records of the Advisory Commission on Intergovernmental Relations (ACIR). As the only site for access to the ACIR electronic records, we frequently received requests for historical publications of the agency from researchers, government administrators, students, and others who found the UNT Libraries' electronic collection by searching the Internet. We proposed to enhance the ACIR electronic collection by making the most important serial titles published by the agency available as electronic documents accessible via the Internet, so in the spring of 1998, I wrote a grant proposal for a pilot project to begin the digitization process. In May of 1998, AMIGOS Bibliographic Council notified me that they were funding the project and then the real work began.

Today, I will be talking about questions you should ask yourself before beginning a digitization project. For each question, options will be discussed and our decisions and processes for the ACIR project will be presented.

Question 1: What Collection Will Be Digitized?

For most of you this question will not be difficult. Being from Texas, the comedian, Jeff Foxworthy, of the "you might be a redneck IF" jokes, is a personal favorite of mine. (Most Texans can connect to "redneck" jokes.) Following his style and my experience with documents librarians, I can confidently say, "If you find you have an uncontrollable urge to digitize everything in your collection and make it all 'permanently publicly available,' you just might be a documents librarian." Your problem will probably be focusing on only one collection as the most likely candidate for digitization. Consider such issues as:

- Uniqueness or rareness of the collection - few Federal documents in our collections are unique, but some may be rare.

- Significance of the collection - many will be an integral part of the history of our country.

- Benefits of electronic access to the information - even if it is not unique or rare, there may be real benefits for electronic access to a collection, including searchability, Internet access, preservation, etc.

Items considered for digitization should also be examined for copyright. This will not be an issue for most collections of government documents. However, if it is an issue, be certain that your copyright information is very current. There are constant changes in the copyright laws, especially with respect to digitizing documents and making them available on the Internet.

Question 2: How Will Funding Be Obtained for the Project?

This question is more difficult to answer than the first. Digitizing any collection requires personnel time (frequently the most expensive element), training, hardware, software, possibly funds for outsourcing parts or all of the project, and many various small expenses. Since funding was not available in my library for a pilot project, I decided to request grant funding. It seemed a reasonable thing to do at the time.

However, if you are writing your first grant proposal, here are a few tips that I learned the hard way. First, be prepared for the significant effort involved in writing the grant proposal and then later writing the grant report. Second, be certain that you check with the grants office on your campus or in the city government before sending out a grant proposal. This will save you many problems over the life of the project. They may add hidden budget costs for various items, such as benefits for project staff. Such costs can affect the amount of money you think you have to spend on the project. They may also ask penetrating questions such as, "If your plan to complete the project is not successful, what is your secondary plan to fulfill the requirements of the grant?" Also, the record keeping must be carefully done so that expenditures are clearly documented.

Question 3: Depending on the Level of Funding Acquired, How Will We Balance Level of Access to the Digitized Documents with Cost of Digitization?

If you found major funding and costs are not an issue, you may want to provide the very best access to thousands of pages of documents by scanning, OCR-ing, marking them up

in HTML, and verifying every word of text for accuracy. We considered this option. However, the costs stated in the project report of the AMIGOS funded study at Oklahoma State University, A Digital Challenge: Bringing Kappler's Indian Affairs: Laws and Treaties to the World Wide Web[1], clearly showed that we could not afford the expensive, time-intensive efforts required to create HTML files from the scanned text, even though we believed that with current technology, HTML files would offer the best Web access.

I stated in my grant proposal that our digitization project would be accomplished by outsourcing high-speed, quantity scanning of approximately 4,200 pages of text. Our goal was to develop a process that balanced level of access with the cost of digitizing and making the data available on the Internet. The objectives included:

A. Providing researchers and the public with electronic access to important publications of the ACIR.
B. Developing a cost-effective process for presenting, on the Internet, extensive collections of PDF files produced from large-quantity, high speed scanning of documents.

Selecting a Vendor

Part of our strategy for controlling costs of scanning included hiring a vendor with the appropriate high-speed scanning equipment to scan the documents. Several vendors were contacted and asked to supply samples of their work. Two vendors agreed to do so and were shipped an issue of an ACIR periodical and a volume of Significant Features of Fiscal Federalism for the test. One vendor supplied us with TIFF files and offered a very low price of 22 cents per page for the black and white scanning. It took several weeks to receive the test scans and several more weeks to retrieve the loaned documents that they scanned. The contact person lacked knowledge about the scanning process and could answer few of my questions about the files.

The other vendor, the Electronic Resource Library Project Lab based at Amarillo College, test scanned the documents and supplied us with TIFF files and with PDF image-plus-text files. The test files and our documents were returned quickly. Their bid was 22 cents per page for TIFF files or 26 cents per page for PDF image-plus-text files. Color scans were offered for document cover pages for an additional 4 cents per page. The director of the lab, Dr. Karen Ruddy, was knowledgeable and prompt with answers to our questions about the files.

The PDF image-plus-text files provided both good quality image files that could be viewed in the free Acrobat Reader and searchable text files. The PDF files were created using the Adobe Capture software, which added the additional benefit of Optical Character Recognition (OCR) to create searchable text. The image file was displayed, but the text file existed and could be searched or copied and pasted. Much of the scanned text was readable by the OCR software.

Pages containing simple text with plain fonts were translated more successfully by the OCR software than non-text material or unusual fonts were. Since the PDF image-plus-text files would allow the additional access of searchable text, they deserved thorough investigation.

We viewed the test files, searched them, copied and pasted from the text, and checked printability. The results exceeded our expectations, so we decided that the PDF image-plus-text files were our best option.

When our experiments showed that the in-house personnel and computer time needed to move TIFF files to PDF image-plus-text files was significant, the decision was made that the extra 4 cents per page charged by the vendor to provide PDF image-plus-text files would be well worth the small extra cost. The PDF image-plus-text files seemed to be our most cost-effective method of balancing issues of access and costs.

Also important influences in our decision to go with the PDF image-plus-text files included:

- Security provided by PDF files.
- Growth in the number of PDF documents offered by Congress and the executive agencies of the Government.
- Health of the Adobe company and the expanding features of the free Acrobat Reader software.

The vendor selected was the Amarillo College Electronic Resource Library Lab. The Lab had previously received grants from the Federal Government to purchase high-speed scanning equipment to digitize documents related to plutonium research. The Lab had also worked with the Department of Energy's "Energy InfoBridge" project, scanning many thousands of pages. Dr. Walter Warnick, director of the Energy Resource Library, highly recommended the Lab. Dr. Ruddy, the Lab director, was interested in outside contracts to keep the Lab personnel and equipment busy. The ACIR scanning project would serve as a pilot project for the Lab to determine if bringing in outside work would be economically feasible.

Question 4: Are All Documents Needed for Digitizing Part of Our Collection, and If Not, How Will They Be Obtained?

This question is particularly relevant if the items are out of print or will be damaged in the scanning process.

My grant proposal stated that approximately 4,200 pages of the most important serial publications of the Advisory Commission on Intergovernmental Relations would be digitized. The ACIR collection at the University of North Texas Libraries was assessed to determine if all issues of our selected serial titles were available in the collection. We estimated that the 1990 -- 1995 volumes of Significant Features of Fiscal Federalism and volume 10 -- volume 20 of Intergovernmental Perspective would approximate the 4,200 pages.

Since high-speed quantity scanning makes use of an automatic paper feeder, any item sent to the Lab would have its binding shaved. The decision was made that retaining a paper copy of each item scanned would be important, so Offers Lists published by the Federal Depository Library Program were monitored frequently to attempt to collect duplicate copies of as many of the publications as possible. When a duplicate could not be located, the

publication would be re-bound after scanning. Duplicates of many of the items were collected when one depository library gave up its depository status and offered all of its collection to other depositories. A few other items were collected at random.

One copy of all volumes of the Significant Features of Fiscal Federalism, except for the 1993 volumes, was in the UNT collection in paper format. The 1993 volumes were in microfiche with the microfiche obviously created from a copy of the original publication. Even though fiche scanners exist, the quality of the fiche copy must be high for the scanned file to be acceptable.

When an inquiry sent to the Texas Library Association Government Documents Round Table listserv showed that all depository libraries had received the 1993 volumes in microfiche, other groups were contacted. The issues were eventually located in the collection of a professor of public administration on the UNT campus. Only a few issues of the periodical, Intergovernmental Perspective, volumes 10 - 12, were missing and were happily supplied by depository librarians at Texas Christian University and the Texas State Library and Archives Commission. As we expanded our scanning back to volume one, other issues were supplied by depository libraries across the country when a request was posted to GOVDOC-L.

Issues or volumes of the titles that were borrowed from individuals or other libraries could not be sent to the Lab to have bindings shaved, so it was determined that these publications would be scanned on an available flat-bed scanner in the UNT Libraries. In July 1998, the first 2,164 pages were shipped to the Lab. An additional 1,436 pages were shipped in August for a total of 3,600 pages. In October, when the UNT Libraries offered additional funding for the project, we shipped an additional 1,872 pages to the Lab. TEXPRESS, the courier system connecting many colleges and universities in Texas, allowed us to ship all documents at no charge to the project.

Question 5: How Will Skilled Personnel Be Found and Training Provided for All Project Participants?

For those of us in academic libraries, students provide a wonderful resource for project personnel. For institutions with library and information science programs, especially skilled graduate students may be found. Our grant proposal included funding to hire a project assistant, so faculty and students who had expressed an interest in the project were notified that we were accepting applications. We were interested in hiring a student who could begin work on the first of August and continue into the fall semester. Interviews were conducted and an extremely well qualified graduate student from the School of Library and Information Sciences was hired.

Training for you as the project manager and for other personnel can be an expensive part of the project. I enrolled in an intensive, three-day class to learn to use the Adobe Acrobat software required to alter and enhance the PDF files. The $450 cost for the class is not an unusual fee and is another cost to include in your grant request. I then instructed the project assistant in the basics of using the software.

The project assistant and I developed a process to create links within the documents,

bookmarks, and other enhancements. Since borrowed items would be scanned on-site, a process for scanning was also created, and the project assistant wrote a procedures manual outlining the process for others to use. As the scanned files were completed by the Lab and sent to us, we enhanced the files by adding bookmarks for the contents of each title, links from the contents pages, and links from indexes when an index was included in the volume. Every page was also checked for readability and printing quality. The procedures manual was edited as needed throughout this process.

Question 6: Do We Have the Technical Skills, Or Access to Qualified Staff, To Solve the Technical Issues of a Digitization Project?

In any digitization project, technical decisions must be made. If you are not fluent in the language of servers, file types, and technical problem solving, be certain that knowledgeable staff are available for consultation.

As we prepared volumes for loading to the Web server, several technical issues required solutions.

- Large file sizes created issues for downloading time.
- Searchable files required a search engine that would index and search PDF files.
- Meta tags had to be defined and entered.
- An overall assessment of the ACIR site was required to integrate the new files effectively.
- Americans with Disabilities Act (ADA) issues were investigated.

File Size

The scanned documents ranged in pagination from approximately 30 pages to over 300 pages. File sizes ranged from 1.8 mega bytes (Mb) to 20 Mb. Downloading such large files over the Web can take considerable time, especially if access is via a modem. When saving the enhanced files, we were careful to use the Acrobat Exchange software's "optimize" function, which helped reduce the size of the files. This, however, did not make the files small enough to have an acceptable download time.

We examined the option of making each page or a few pages into separate files, then creating some type of navigation system to allow the user to move on to the next file (next page of the document). We visited two sites that use this method. Even though it did reduce download time, we felt it was cumbersome for the user, and it would increase our costs significantly by requiring additional time to prepare our files for the Web.

Searching for other options, we discovered in a mailing list archive called the "PDF Archive," a possible solution called "byteserving."[2] It involved setting up the files correctly and having Web server software that supports the "Byte Range Retrieval Extension to HTTP" protocol. This server software has the capability to "serve" to the user only one page

at a time of a PDF file. This method requires the user to change only one setting in the Acrobat Reader Preferences to disallow "background downloading."

The user can then move through the document using the Acrobat Reader's functions or the links and bookmarks we created. Since the UNT Libraries' Web server already had one of the software packages that supports byteserving, we tested this option and decided it would be the best option for us. On our Web interface page, we asked the user to link to another page to find out about "Faster Downloads."[3] There we explained byteserving and how "preferences" in the Acrobat Reader could be easily altered for faster downloading of the files.

Searching PDF Files

From the beginning of the project, our goal to make the ACIR Web site searchable was an important part of maximizing access to the digitized collection. We quickly learned that many of the well-known search engines would not index and search PDF files. We spent a considerable amount of time viewing and reading about our options. The project assistant created a table outlining our most promising options. Infoseek's Ultraseek Server, Microsoft Index Server, and Verity Search were our best options.

Infoseek was reasonably priced, had automatic re-indexing, supported sophisticated search queries and responses, and was Y2K compliant. Microsoft Index Server was free with our Windows NT 4.0 server software and had automatic re-indexing. However, it did not rank search results or detect duplication, and it often included HTML characters when creating summaries. The Microsoft Index Server did offer a PDF filter that could be installed so that PDF files would be indexed. The Verity search engine provided a special filter to search over 200 file formats and used Meta tags to control summaries, so responses to a query were controlled by the metadata entered for each PDF file. Our investigation also revealed that the Netscape Compass Server used the Verity search engine software.

Since the University used the Netscape Compass Server without cost, it was our best option. It required the addition of a PDF filter for indexing PDF documents. However, Netscape recently made the announcement that educational institutions would no longer have free access to their Netscape Compass Server software. It is unclear at this time how this will affect installation of the software. There will undoubtedly be problems to solve as we activate it or our second choice, the Microsoft Index Server, and create the appropriate CGI scripts.

Search engines examined but rejected for various reasons included:

- SWISH-E, because it runs on a Unix platform and we run Windows NT.
- Excite for Web Servers, because it does not search PDF files.
- Harvest, because it also runs on a Unix platform and the project appeared to be bankrupt.
- Sage (NUD*IST 4), because it is a project-based search for personal computers.

- Excalibur RetrievalWare, although it has many wonderful features, because it is very expensive and really more than we needed.

Metadata

Metadata are used to describe an information resource. Whatever the file type used in a digitization project, Meta tags are important for accurate retrieval of documents. The Acrobat Exchange software allows entry of four Meta tags for each PDF file created.

The Meta tags are very important because this is the information used to build the index list when searching PDF files. Without Meta tags, the index list often contains the URL as the title of the document and the first few words of readable text in the document as the description. Such a list may not be an accurate description of the document, and if the OCR software was unable to read the first few words, the information may even be unreadable. For this reason, the decision was made to include metadata for every PDF document.

Much of the data entry for the Meta tags is awaiting the activation of the search engine. Until we see how the selected search engine builds the indexes and displays the index lists, we cannot know what information to enter on each line of the Meta tags.

All accompanying HTML pages were created with title, keyword, and description Meta tags. Our search of the literature found articles outlining research that showed HTML documents with title, keyword, and description Meta tags were ranked higher on index lists built by some Web search engines. Also, most Web search engines use the title and description Meta tags to build the index list. When the Meta tags are not present, the title displayed is often either the URL of the page or "No Title," and the first few words of the document become the description.[4]

Integrating Digitized Files Into a Web Site

Whatever file type chosen for a digitization project, Web access must be provided in an organized and varied manner. As librarians, our skills as organizers of information certainly assist with this part of the project. Support pages that may be required include pages for browsing by topic or by title, bibliographies, help pages, or pages with historical or related information. Specialists or experts may be consulted for input for this part of the project.

Realizing that the hyperlink properties of HTML documents could assist us with offering multiple access points to the full-text PDF files, we examined the overall design of the ACIR Web site. Already contained on the site were the electronic files of the ACIR as they appeared when the ACIR closed in 1996, and the UNT Libraries agreed to provide permanent public access to the files. This part of the Web site could not be altered from the way the files appeared when the agency closed.

To enhance the original files, we added a brief history of the ACIR and a bibliography of the publications of the ACIR. Relevant dates and citations for laws that created or affected the ACIR were collected for the history of the agency, and the bibliography of ACIR publications was compiled and added.

Additional Web pages created to provide access to the PDF documents and to provide technical information about the site included:

- A "Browse Titles" page to allow access to the PDF documents from an alphabetical listing of titles <**http://www.library.unt.edu/gpo/acir/browsetitles.html**>
    - A page to explain byteserving and to describe the simple steps to allow for faster downloads of the files <**http://www.library.unt.edu/gpo/acir/technicaldoc.htm**>

      Users without the Acrobat Reader software were linked to the Adobe Web site so that they could download the free software.

      In addition to offering a searchable site as discussed above, we decided to make the bibliography an important additional access point to the full-text documents. To improve access, the organization of the bibliography needed to follow the way researchers and experts in the field of intergovernmental relations search for information. The project assistant examined a bibliography published by the ACIR in the periodical, Intergovernmental Perspective,[5] and then made suggestions for the reorganization of our Bibliography of the Publications of the ACIR page.

      Intergovernmental relations expert and Assistant Professor of Public Administration at the University of North Texas, Dr. Michael McGuire, reviewed the suggestions for reorganizing the bibliography. Working with his comments we created a final plan for the reorganization. Each title in the bibliography will link to the full-text document as it is added to the site and to the corresponding MARC record in the UNT Libraries' online catalog to provide additional metadata about each title. In time MARC records in the Libraries' catalog will link to the electronic full-text document.

      ADA Concerns

      Access to the PDF files for persons with impared vision was a concern that required considerable research. The project assistant discovered that the text readers commonly used by persons with visual impairments would not read PDF files. He also discovered that Adobe offered a free program that would translate PDF files into HTML files which text readers can translate.

      T.V. Raman, a senior computer scientist at Adobe Systems, created the program, called Acrobat Access.[6] Mr. Raman is blind and created the program for his own use. The computer program works best on documents composed of simple text, since graphics and other visually rich features do not translate well. The free software was available for downloading from the Adobe site. However, when we tried to download and install it, a number of problems arose. We discovered that the program was designed to run in Windows 3.1 and had not been updated since 1996. There were some conflicts with running it in Windows 95.

Even though the problems were not insurmountable, a user would have to be technologically skilled to install and operate the program effectively. Adobe offered for sale for about $100 another software package called Genus HTML. It is compatible with Windows 95, Windows NT, and Macintosh operating systems and translates PDF files to HTML. Like the Acrobat Access software, Genus HTML works best with PDF files containing simple text.[7]

When the free Acrobat Access software was tested, the PDF image-plus-text files on our site were readable by the visually impaired text readers, but only the text that could be captured by the OCR software was available to the user. This meant that words not read by the software in the OCR process were not readable by the text reader programs. To fix this problem, every word of the text would require verification and correction by a person, which would require many hours of work at high costs. The purpose of this project was to develop a method of digitizing a collection that would balance access and costs. At this point, we have not discovered a solution for this problem.

As you can see from these examples, it is imperative that project staff possess excellent technical skills.

Question 7: Can a Digitization Project Be Accomplished at a Reasonable, Predictable Cost?

Hoping to answer this question, we kept statistical data about each step in the process during the first four months of the project. Times were logged for downloading the files from the Lab's server and for each step in the enhancement of the files. Later, as we scanned in-house several issues of the periodical title, times were noted for both the scanning and the OCR process.

The first 3,600 pages scanned by the vendor cost $938.44. This included 3,539 black and white scans and 61 color scans. Nine volumes of Significant Features of Fiscal Federalism accounted for 2,164 pages, with thirty-seven issues of the periodical, Intergovernmental Perspectives, accounting for the remaining 1,436 pages. (See Figure 1 and Figure 2) Enhancements for each volume/issue included creation of bookmarks, creation of links from the "contents" pages, and, for Significant Features of Fiscal Federalism, creation of links from the index. (See Figure 7) For a graphical display of the data, see Figures 3, 4, and 8.

As the in-house work progressed on the files, we noticed a significant increase in the speed of completion of each title. The project assistant became very proficient with the Acrobat Exchange software. The first two volumes of Significant Features of Fiscal Federalism required an average of about 240 minutes each to complete compared with about 125 minutes each to complete the last two volumes, even though the files sizes were similar. (Figure 9) Later in the project, the speed of completion became static, with each periodical issue requiring about 28 minutes. The average time and costs

per page were:

No. of Pages Scanned	In-house Time Per Page	Scanning Cost Per Page	Total Cost Per Page
1,432 pages of Inter-governmental Perspective scanned by vendor	.74 minutes per page @ $10 per hour = .12 per page	.26 per page	.38 per page
356 pages of Inter-governmental Perspective scanned in-house	3.23 minutes per page @ $10 per hour = .54 per page	N/A	.54 per page
2,188 pages of SFFF scanned by vendor	.85 minutes per page @ $10 per hour = .14 per page	.26 per page	.40 per page
Average minutes per page	1.61 minutes per page	**Average cost per page**	.44 per page

The data for in-house time spent with Significant Features of Fiscal Federalism files (SFFF) indicates that the average cost per page would be lower if the first two documents were excluded. As proficiency of the project assistant with the software increased, the amount of time required to complete a document decreased considerably. This is clearly illustrated by Fig. 9. After completion of the SFFF files, the periodical files were enhanced with little variation in time per page. (Figures 5 and 6) Figure 6 also illustrates the widely varying time required for in-house scanning of Intergovernmental Perspective. This time variation can be credited to several factors including:

1. Experience of the person scanning documents.
2. Speed of the scanner.
3. Speed of the computer and available memory in the computer.
4. Speed of moving the data over the network.
5. Condition of the document to be scanned.

With training and with the purchase of suitable hardware, most of the factors can be overcome. However, the condition of the document cannot be predicted. Some documents have detailed graphs; others have color print on a differently colored background. Some have very small font size; others may have unreadable text. The condition of the document brings unpredictability for scanning times.

The 15-cents-per-page difference in the in-house scanning cost and the out-sourced cost is significant only when a large project is undertaken. For a

100,000-page project, outsourcing would result in a significant savings of $15,000. Also, when outsourcing the scanning, the vendor would deal with the problems encountered with the condition of each document. However, for small projects, in-house scanning is a reasonable alternative.

As the project manager, my hours are difficult to calculate because much of the management of the project was integrated with the other tasks common to our academic librarian's day. I supervised the project assistant, facilitated the workflow, and coordinated with the Libraries' LAN/PC Management department. Additional hours were spent with problem solving and attempting to "see the big picture," e.g., determining how all the pieces would fit together to create the Web site. Since projects that increase knowledge and growth as a professional are considered an important part of a UNT librarian's activities, this project was included as part of the workload.

According to the project report from the AMIGOS funded study, A Digital Challenge: Bringing Kappler's Indian Affairs: Laws and Treaties to the World Wide Web,[1] pages of text were scanned, then the text was verified and marked up in HTML requiring an average of 66 minutes per page of staff time plus an average of 32.5 minutes per page of student assistant time. Comparison with our average total of 1.61 minutes per page makes it clear that the use of vendor scanning and PDF image-plus-text files can significantly reduce the cost of digitizing a collection while still providing good access. Of course, creation of PDF image-plus-text files by high-speed vendor scanning is a production digitization method that is not appropriate for all types of documents, but when appropriate, the cost savings are notable.

Current Status of the ACIR Project

Now available on the Web site are more than 7,000 pages of ACIR documents. Excluding the initial training, our average cost per page was 44 cents. In the short term, we continue to add individual ACIR documents as requests for specific titles are received. Currently available are:

Intergovernmental Perspectives

1975: Vol. 1 Issue 1

1976: Vol. 2 Issue 1 Issue 2 Issue 3 Issue 4

1977: Vol. 3 Issue 1 Issue 2 Issue 3 Issue 4

1978: Vol. 4 Issue 1 Issue 2 Issue 3 Issue 4

1979: Vol. 5 Issue 1 Issue 2 Issue 3 Issue 4

1980: Vol. 6 Issue 1 Issue 2 Issue 3 Issue 4

1981: Vol. 7 Issue 1 Issue 2 Issue 3 Issue 4

1982: Vol. 8 Issue 1 Issue 2 Issue 3 Issue 4 (Included with Vol. 9, Issue 1)

1983: Vol. 9 Issue 1 Issue 2 Issue 3 Issue 4

1984: Vol. 10 Issue 1 Issue 2 Issue 3 Issue 4

1985: Vol. 11 Issue 1 Issue 2-3 Issue 4

1986: Vol. 12 Issue 1 & 2 Issue 3

1987: Vol. 13 Issue 1 Issue 2 Issue 3 & 4

1988: Vol. 14 Issue 1 Issue 2 Issue 3 Issue 4

1989: Vol. 15 Issue 1 Issue 2 Issue 3 Issue 4 (Includes bibliography of ACIR Publications-- 1961-1989)

1990: Vol. 16 Issue 1 Issue 2 Issue 3 Issue 4

1991: Vol. 17 Issue 1 Issue 2 Issue 3 Issue 4

1992: Vol. 18 Issue 1 Issue 2 Issue 3 Issue 4

1993: Vol. 19 Issue 1 Issue 2 Issue 3

1994: Vol. 20 Issue 1 Issue 2 Issue 3 (Includes bibliography of ACIR Publications - 1960 - 1994)

Significant Features of Fiscal Federalism

1979 - 1980

1980 - 1981

1981 - 1982

1982 - 1983

1984

1985 - 1986

1988 Vol. 1

1989 Vol. 1 Vol. 2

1990 Vol. 1 Vol. 2

1991 Vol. 1 Vol. 2

1992 Vol. 1 Vol. 2

1993 Vol. 1 Vol. 2

1994 Vol. 1 Vol. 2

1995 Vol. 1 Vol. 2

The expertise we gained during the project is also being put to good use. Within the UNT Libraries, the processes developed by our project were shared with the Music Library's staff. They digitized eighteen volumes of Jean-Baptiste Lully's scores using the method we developed for in-house scanning and processing of files. Their project was funded by a TEXSHARE grant and completed quickly using our process. Our technical assistance support page was also adapted and used for their project.

Future plans for the project

Using the results of this pilot project, UNT Assistant Professor and specialist in intergovernmental relations, Dr. Michael McGuire, and I are writing a grant proposal to obtain funding to complete the digitization of the ACIR publications. We are communicating with other scholars and researchers in the field for advice and assistance and for the purpose of publicizing the collection. If funding becomes available, approximately 60,000 additional pages will be digitized and made available to the public.

Notes:

6. See the report at:
   <**www.library.okstate.edu/kappler/intro.htm**>
7. Read about "byteserving" at:
   <**www.adobe.com/prodindex/acrobat/byteserve.html**>
8. View the Technical Assistance page at:

   <**www.library.unt.edu/gpo/acir/technicaldoc.htm**>

9. See the Web site: Search Engine Features Chart
   <**www.searchenginewatch.com/webmasters/features.html**>

See article:

Turner, Thomas P. and Lise Brackbill. "Rising to the Top: Evaluating the Use of the HTML Meta Tag to Improve Retrieval of World Wide Web Documents Through Internet Search Engines," Library Resources and Technical Services, Vol. 42 (4), 1998, p. 258 - 271.

10. See the Bibliography at: <**www.library.unt.edu/gpo/acir/periodical/ipsfv20n4. pdf**>
11. To read about the Acrobat Access software, go to: <**www.adobe.com/supportservice/custsupport/LIBRARY/3b7e.htm**>
12. To read about Adobe's Genus HTML software, go to: <**www.pluginsource.com/acrobat/genushtml.html**>

Outsourced Scanning of Intergovernmental Perspectives										
Title	**Down-loading**	**Bookmark and Page Setup**	**Contents Linked**	**Cover Touch-Up**	**Saving**	**Total**	**File Size (M)**	**Min/M**	**Pages**	**Min/Pg**
IPSFv10n184	17	6	3	2	2	30	3.52	8.52	40	0.75
IPSFv10n284	18	7	3	1	2	31	2.81	11.03	36	0.86
IPSFv10n384	16	8	5	1	2	32	2.76	11.59	32	1.00
IPSFv11n185	14	4	6	2	2	28	4.40	6.36	48	0.58
IPSFv11n2n385	19	5	5	1	2	32	3.03	10.56	24	1.33
IPSFv11n485	21	6	7	2	2	38	3.55	10.70	40	0.95
IPSFv14n188	8	3	2	X	2	15	2.70	5.56	28	0.54
IPSFv14n288	5	4	3	3	2	17	2.12	8.02	24	0.71
IPSFv14n388	7	5	3	2	2	19	2.35	8.09	28	0.68
IPSFv14n488	6	4	3	2	2	17	2.44	6.97	28	0.61
IPSFv15n189	8	3	2	3	2	18	3.48	5.17	40	0.45
IPSFv15n289	13	4	2	2	2	23	3.04	7.57	36	0.64
IPSFv15n3	10	4	2	2	2	20	3.64	5.49	40	0.50

IPSFv15n4	12	5	5	1	4	27	3.80	7.11	40	0.68
IPSFv16n1	10	6	5	X	2	23	3.28	7.01	40	0.58
IPSFv16n2	10	4	5	X	2	21	4.07	5.16	48	0.44
IPSFv16n3	15	6	6	X	2	29	3.87	7.49	32	0.91
IPSFv16n4	20	7	6	X	2	35	3.36	10.42	40	0.88
IPSFv17n1	8	6	5	X	2	21	4.34	4.84	52	0.40
IPSFv17n2	6	8	7	X	2	23	3.18	7.23	40	0.58
IPSFv17n3	8	8	7	X	2	25	3.95	6.33	48	0.52
IPSFv17n4	20	8	6	2	2	38	4.98	7.63	60	0.63
IPSFv18n1	20	6	4	10	2	42	4.06	10.34	48	0.88
IPSFv18n2	20	6	4	2	2	34	3.40	10.00	40	0.85
IPSFv18n3	20	6	7	6	2	41	2.80	14.64	32	1.28
IPSFv18n4	22	4	3	8	2	39	2.86	13.64	36	1.08
IPSFv19n1	15	8	4	6	2	35	3.57	9.80	40	0.88
IPSFv19n2	12	8	6	8	2	36	3.70	9.73	44	0.82
IPSFv19n3	17	8	5	X	2	32	3.30	9.70	36	0.89
IPSFv20n1	18	4	2	1	2	27	4.38	6.16	48	0.56
IPSFv20n2	23	4	3	6	2	38	3.05	12.46	40	0.95
IPSFv20n3	21	2	2	3	2	30	3.14	9.55	44	0.68
IPSFv12n1n2	15	4	3	1	2	25	3.03	8.25	36	0.69
IPSFv12n3	12	4	2	2	2	22	2.82	7.80	32	0.69
IPSFv13n1	12	3	2	2	2	21	3.07	6.84	40	0.53
IPSFv13n2	14	4	3	4	2	27	6.19	4.36	32	0.84
IPSFv13n3n4	18	4	2	2	2	28	3.20	8.75	40	0.70
Total	530	196	150	21	76	1039	127.24		1432	
AVG	14.32	5.30	4.05	2.35	2.05	28.08	3.44	8.40	38.70	0.74

Figure 1

In-House Scanning of Intergovernmental Perspectives											
Title	**Scanning**	**OCR**	**Bookmark and Page Setup**	**Contents Linked**	**Cover Touch-Up**	**Saving**	**Total**	**File Size (M)**	**Min/M**	**Pages**	**Min/Pg**
IPSFv10n484	75	40	15	2	1	4	137	5.01	27.35	32	4.28
IPSFv9n183	95	45	9	3	3	3	158	13.20	11.97	48	3.29
IPSFv9n283	40	45	8	3	2	2	100	11.10	9.01	36	2.78
IPSFv9n383	25	25	8	2	1	2	63	6.63	9.50	32	1.97
IPSFv9n483	72	35	4	2	1	2	116	3.51	33.05	32	3.63
IPSFv8n182	55	62	4	2	1	2	126	4.64	27.16	44	2.86
IPSFv8n282	60	72	6	4	2	2	146	3.69	39.57	36	4.06
IPSFv8n382	72	86	4	3	1	2	168	3.01	55.81	32	5.25
IPSFv7n181	30	30	4	2	1	2	69	4.75	14.53	32	2.16
IPSFv7n281	40	50	4	3	4	3	64	3.24	19.75	32	2.00
Total	**564**	**490**	**66**	**26**	**17**	**24**	**1147**	**58.78**		**356**	
Average	**56.40**	**49.00**	**6.60**	**2.60**	**1.70**	**2.40**	**114.70**	**5.88**	**24.77**	**35.60**	**3.23**

Figure 2

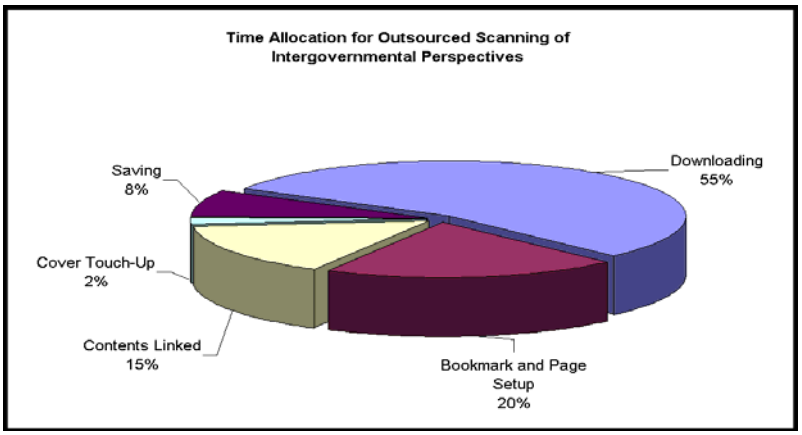

Figure 3

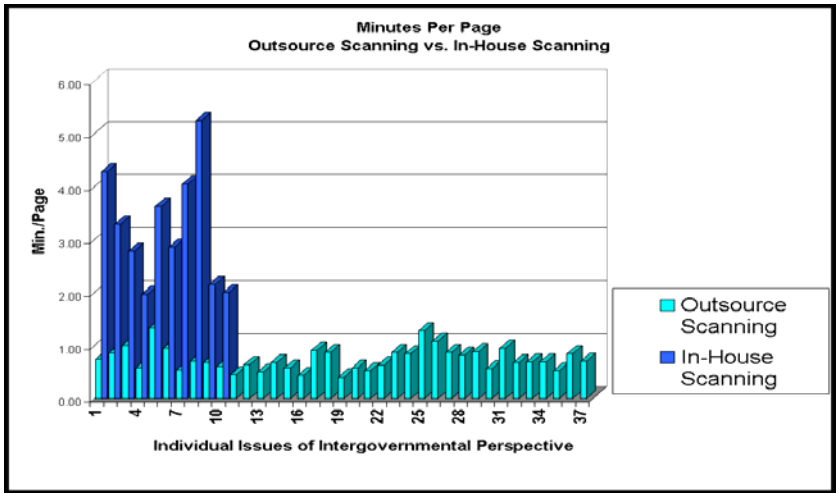

Figure 4

	Down-load Time	Bookmark and Page Setup	Contents Linked	Index Linked	Saving	Total	File Size (M)	Min/M	Pages	Min/Pg
SFFF 1991 V. 1	15	90	45	135	5	290	10.04	28.88	188	1.54
SFFF 1990 V. 1	19	112	25	30	5	191	7.64	25.00	164	1.16
SFFF 1990 V. 2	25	35	65	45	5	175	13.38	13.08	284	0.62
SFFF 1992 V. 1	14	100	30	45	5	194	9.46	20.51	196	0.99
SFFF 1992 V. 2	30	105	55	190	5	385	16.87	22.82	372	1.03
SFFF 1991 V. 2	23	120	30	115	5	293	16.25	18.03	344	0.85
SFFF 1994 V. 1	12	25	10	26	5	78	8.23	9.48	180	0.43
SFFF 1994 V. 2	25	30	43	78	5	181	12.25	14.78	284	0.64
SFFF 1995 V. 1	16	20	8	19	5	68	8.03	8.47	176	0.39
Total	179	637	311	683	45	1855	102.15		2188	
AVG	19.89	70.78	34.56	75.89	5.00	206.11	11.35	18.16	243.11	0.85

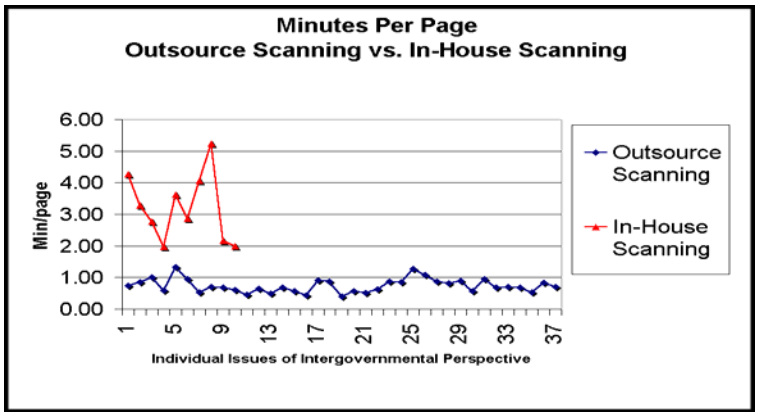

Figure 5

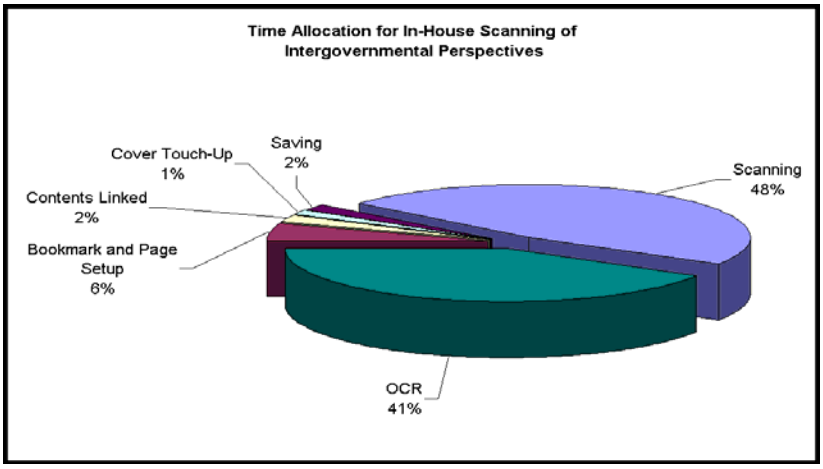

Figure 6

Significant Features of Fiscal Federalism

Time Allocation Statistics

Figure 7

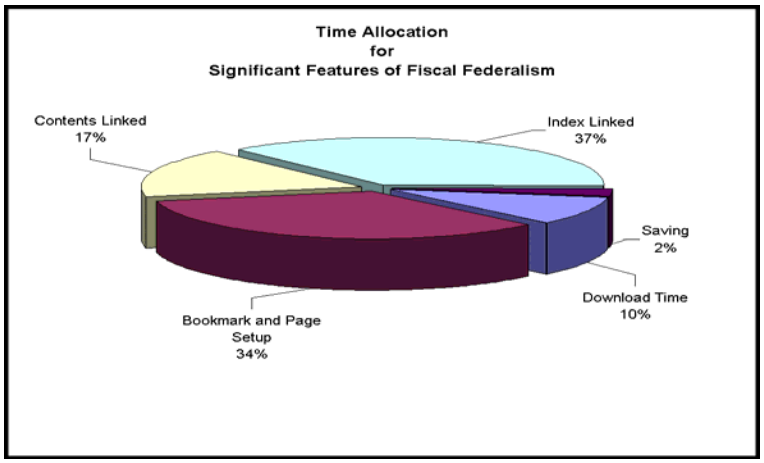

Figure 8

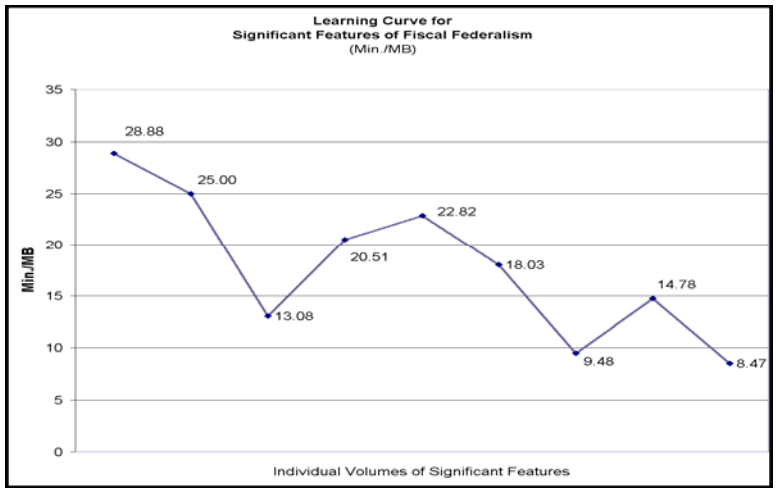

Figure 9