

**Digitizing Historical Publications:  
Enhancing the Official Electronic Collection of the Advisory Commission on  
Intergovernmental Relations**

**Amigos Fellowship Final Report  
April 1999**

**Submitted By**

**Cathy Nelson Hartman  
University of North Texas Library**

## **Introduction**

In May of 1998, AMIGOS Bibliographic Council funded a project to digitize serial titles published by the Advisory Commission on Intergovernmental Relations and make them available on the World Wide Web. Our goal was to develop a process that balanced the level of access with the cost of digitizing documents and making them available on the World Wide Web. The abstract describing the project and the objectives proposed for the project are outlined in the following excerpts from the grant proposal.

### **Abstract:**

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 official and the only site for ACIR electronic records, requests for historical publications of the agency are frequently received from researchers, government administrators, students, and others who find UNT Libraries electronic collection by searching the Internet. We propose to enhance the current ACIR electronic collection by making the most important serial titles published by the agency available as electronic documents accessible via the Internet. The serial titles are no longer in print and, to our knowledge, are not available in electronic format. This digitization project will be accomplished by outsourcing high-speed, quantity scanning of approximately 4,200 pages of text. The high speed scanning will provide us with “.pdf” files. We will then develop a method for organizing and presenting the files on the World Wide Web designed to provide broad access using the free Acrobat Reader. Our goal will be to develop a process that balances level of access with the cost of digitizing and making the data available on the Internet.

### **Objectives of the proposal:**

- A. Provide researchers and the public With electronic access to important publications of the ACIR.
- B. Develop a cost-effective process for presenting, on the Internet, extensive collections of .pdf files produced from large-quantity , high speed scanning of documents.
- C. Enhance the knowledge and project management skills of the applicant relative to digitizing and making available on the Web important government documents.
- D. Share knowledge gained from the project with depository librarians and others interested in quantity digitization of historical information at an affordable cost.

## **Beginning the Project**

In the initial phase of the project, several issues were addressed before the actual digitization process could begin.

- **A vendor was selected for the high-speed scanning of the documents.**
- **Documents were collected.**
- **Training was provided for the project manager, and a graduate student assistant was hired and trained.**
- **Selecting a Vendor**

Part of the 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 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 periodical 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, Electronic Resource Library Project Lab based at Amarillo College, test scanned the periodical and supplied us with TIFF files and with PDF image-plus-text files. The test results 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. Since the PDF image-plus-text files would allow the additional access of searchable text, they were investigated thoroughly. 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 for the PDF image-plus-text files would be well worth the small extra cost. The PDF image-plus-text files were clearly our most cost-effective method of balancing issues of access and costs.

Another option considered was creating HTML files. However, the costs stated in the project report of the AMIGOS funded study, *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 HTML files would offer the best Web access.

The vendor selected was the Amarillo College Electronic Resource Library Lab. The Lab 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 the outside work would be economically feasible.

- **Collecting the Documents**

The 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 rebound. Duplicates of many of the items were collected when one depository library gave up its depository status and was required to offer all of its collection to other depositories. A few other items were collected at random.

All volumes of the *Significant Features of Fiscal Federalism*, except for the 1993 volumes, were in the UNT collection in paper format. The 1993 volumes were in microfiche. 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.

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 vendor. 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, another 1,872 pages were shipped to the Lab. All documents were Hartman. AMIGOS Fellowship Program Report, 03/27/99, page 4 shipped at no charge to the project via TEXPRESS, the courier system connecting many colleges and universities in Texas.

#### • **Training and Personnel**

The grant proposal included funding to hire a project assistant for the month of August, 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 with funding from a Graduate Library Assistantship program. Interviews were held and an extremely well-qualified graduate student from the School of Library and Information Sciences was hired.

The project manager enrolled in an intensive, three-day class to learn to use the Adobe Acrobat software required to alter and enhance the **PDF** files. The UNT Libraries generously underwrote the \$450 cost for the class. The project manager then instructed the project assistant in the basics of using the software, and a process was developed 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. (Appendix A)

#### • **Technical Issues**

As the scanned files were completed by the Lab and sent to us, we enhanced the files by adding bookmarks for the contents of the work, 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.

As volumes were prepared for loading to the Web server, there were several points that 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.

- Web pages had to be designed and created to serve as an interface for the PDF files and provide technical assistance.
- 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. To download 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, and even though it does reduce download time, we felt it was cumbersome for the user.

Searching for other options, we discovered in the "PDF Archive" mailing list archives a possible solution called "byteserving."<sup>(2)</sup> 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 only requires the user to change 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,"<sup>(3)</sup> and there we explained how the preferences in the Acrobat Reader could be 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. (Appendix B) 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 was based on the Verity search engine. 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. At this writing, we are waiting for the search software to be installed by our Libraries' LAN/PC Management Department. 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 rejected for various reasons were SWISH-E, Excite for Web Servers, Harvest, Sage (NUD\*IST 4), Excalibur RetrievalWare. To read specific reasons for the rejection of these options, please see Appendix B.

- **Metadata**

Metadata is used to describe an information resource, and the Acrobat Exchange software allows entry of four MET A tags for each document created. The MET A 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 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 MET A tags is awaiting the activation of the search engine. Until we see how the search engine builds the indexes and the index lists, we cannot know what information to enter on each line of the MET A tags.

All accompanying HTML pages were created with title, keyword, and description MET A tags since our research showed that HTML documents with these MET A tags were ranked higher on index lists by some Web search engines. Also, most Web search engines use the title and description MET A tags to build the index list. When the MET A 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)

- **Web Pages to Support the Full- Text Files**

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://NW.N.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://NW.N.I 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.

- **Integrating PDF Files Into the ACIR Web Site**

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. Contained on the site were the electronic files of the ACIR as they appeared when the ACIR closed in 1996. Through a 1997 agreement with the Government Printing Office, the UNT Libraries 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
- A Bibliography of the Publications of the ACIR .
- Browse Titles of Full- Text Documents

Relevant dates and citations for laws that created or affected the ACIR were compiled into a brief history of the agency and added to the Web site. (<http://WNW.library.unt.edu/gpo/acir/history.html>).

In addition to offering a searchable site as discussed above, we decided to create a bibliography as an important additional access point to the full-text documents. 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 and to the corresponding MARC record in the UNT Libraries online catalog to provide additional metadata about each title.

A “Browse Titles” Web page allowed access to the PDF documents from an alphabetical listing of titles. (<http://www.library.unt.edu/gpo/acir/browsetitles.html>) Byteserving and the simple steps to allow faster downloads of files were described in another page (<http://www.library.unt.edu/gpo/acir/technicaldoc.htm>), and users without the Acrobat Reader software were linked to the Adobe Web site for the free software.

- **ADA Concerns**

Access to the PDF files for persons with limited 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. 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 is 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.

Another software package was discovered called Genus HTML. Offered for sale (about \$100) by Adobe, it is compatible with Windows 95, Windows NT, and Macintosh operating systems. It translates PDF files to HTML and, like the Acrobat Access software, works best with PDF files containing simple text. (7)

When the Adobe Access software was used, the **PDF** image +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.

## Cost analysis

For the first four months of the project, we kept statistical data about each step in the process. 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 (See Appendix C for the Scanning Log). 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. 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. (For detailed data, see Appendix D)

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. 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 Intergovernmental Perspective scanned by vendor | .74 minutes per page @ \$10 per hour = .12 per page  | .26 per page                    | .38 per page        |
| 356 pages of Intergovernmental 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        |
| <b>Average minutes per page</b>                                | 1.61 minutes per page                                | <b>Average minutes per page</b> | .44 per page        |

The data for in-house time spent with *Significant Features of Fiscal Federalism* files (SFFF) (Appendix D, Fig. 6) 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. 7 in Appendix D. After completion of the SFFF files, the periodical files were enhanced with little variation in time per page. (See Appendix D, Fig. 5) Fig.5 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.

The project manager's 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. The project manager 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, edited, and marked up in HTML requiring an average of 66 minutes per page of staff time and 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 vendor scanning and PDF image - plus-text files can significantly reduce the cost of digitizing a collection without significantly sacrificing access. Of course, the PDF image-plus-text files provided 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 is notable.

### Analysis of objectives

In the introduction, the four objectives of the project were listed as they appeared in the original grant proposal. Did we accomplish our objectives?

**Objective 1:** Provide researchers and the public with electronic access to important publications of the ACIR.

Now available on the ACIR Web site, are the following volumes and issues of the two most important serial titles published by the ACIR.

#### *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: Volume 1  
 1989: Volume 1 | Volume 2  
 1990: Volume 1 | Volume 2  
 1991: Volume 1 | Volume 2  
 1992: Volume 1 | Volume 2  
 1993: Volume 1 | Volume 2  
 1994: Volume 1 | Volume 2  
 1995: Volume 1 | Volume 2

**Objective 1:** Provide researchers and the public with electronic access to important publications of the ACIR.

Now available on the ACIR Web site, are the fo

**Objective 2:** Develop a cost-effective process for presenting, on the Internet, extensive collections of pdf files produced from large-quantity, high speed scanning of documents.

With 6,843 pages now available on the ACIR Web site and with an average cost per page of 44 cents, we conclude that the process developed by this project is cost effective.

**Objective 3:** Enhance the knowledge and project management skills of the applicant relative to digitizing and making available on the Web important government documents.

The knowledge and skills of the project manager were significantly enhanced. Benefits included learning to:

- Monitor and coordinate the budget with the University's Grants office. .Resolve hiring and training issues.
- Solve problems with files, software, and hardware involved in digitization.
- Coordinate the project with other departments to arrange for computer time, server space, software loading, and problem solving.
- Share knowledge gained with other departments in the **UNT** Libraries.

**Objective 4:** Share knowledge gained from the project with depository librarians and others interested in quantity digitization of historical information at an affordable cost.

In addition to this report, the project manager will present a paper at the Depository Library Conference in Washington, D.C., in April 1999, based on knowledge gained from the project. The presentation will be published in the proceedings of the Conference, usually published about four months after the Conference. Also, an article is planned for later this year, and other opportunities for presenting to interested groups will be sought.

Within the UNT Libraries, the processes developed by our project were shared with the Music Department. Staff in the Music Department 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. The files are available on the Web. (<http://IIVW.library.unt.edu/projects/lully/lullyhom.html>) Our technical assistance support page was also adapted and used for their project.

| Projected        | Actual           | Task   |
|------------------|------------------|--|
| 1998 May to July | 1998 May to July | Collect titles to be scanned and send to vendor for s  |
| 1998 August      | 1998 August      | Hire graduate student assistant and begin experime investigation of a cost-effective process of organizi making the files available on the Internet. |
| 1999 June        | 1998 September   | Have available for access on the ACIR Web site, Sig Features of Fiscal Federalism, Volume 1, 1995.   |
| 2000 January     | 1998 November    | Have available for access on the ACIR Web site, Intergovernmental Perspective, Volumes 16 – 20.  |
| 2000 March       | 1999 March       | Final Report sent to AMIGOS.   |

### Future Plans for the project

UNT Assistant Professor and expert in intergovernmental relations, Dr. Michael McGuire, is working with us to write 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. In the short term, individual documents are added as requests for specific titles are received.

## Notes:

(1) See the report at:

<<http://www.library.okstate.edu/kappler/intro.htm>>

(2) Read about "byteserving" at:

<<http://www.adobe.com/prodindex/lacroba/byteserve.html>>

(3) View the Technical Assistance page at:

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

(4) See the Web site:

Search Engine Features Chart

<<http://www.searchenginewatch.com/webmasters/features.html>>

See article:

Turner, Thomas P. and Lise Brackbillo "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, pp 258-271

(5) See the Bibliography at:

<<http://www.library.unt.edu/gpo/acir/periodical/ipsfv20n4.pdf>>

(6) To read about the Acrobat Access software, go to:

<<http://www.adobe.com/supportservice/custsupport/LIBRARY13b7e.htm>>

(7) To read about Adobe's Genus HTML software, go to: <<http://www.pluginsource.com/acrobat/genushtml.html>>

## Attachments:

Appendix A: Project Manual

Appendix B: Search Engines

Appendix C: Scanning Log

Appendix D: ACIR Project Data

## Project Participants

Cathy Hartman -Applicant and project manager. Documents Librarian: Electronic Resources Coordinator, University of North Texas Libraries.

Involved in every aspect of the project.

Luke Griffin -Project Assistant. Graduate Student in the UNT School of Library and Information Sciences and Graduate Library Assistant in the UNT Libraries,

Involved with development of the method for enhancement and manipulation of the files and with the scanning process procedures. Involved with solving many of the early technical issues, such as file size, search engines, and Web interface. Accomplished all of the production work for the project.

Karen Neal -Graduate Student in the UNT School of Library and Information Sciences and Graduate Library Assistant in the UNT Libraries.

Involved with implementing the process defined in the initial part of the project to make the files available on the Web and with the ACIR Web site's overall assessment.

Project Advisors: Melody Kelly -Head, Government Documents Department, UNT Libraries and Robert Pierce -Head, LAN/PC Management Department, UNT Libraries

# **Appendix A**

## **ACIR Project PDF Manual**

## Contents

|                               |   |
|-------------------------------|---|
| DirectoryStructure . . . . .  | 1 |
| Personal Directory . . . . .  | 1 |
| Shared Directory . . . . .    | 1 |
| PDF Files . . . . .           | 1 |
| File Size . . . . .           | 1 |
| ByteServing . . . . .         | 1 |
| Document Processing . . . . . | 2 |
| Process I . . . . .           | 2 |
| Process II . . . . .          | 3 |
| Cover Page . . . . .          | 4 |
| Other Pages . . . . .         | 4 |
| ADA Compliance . . . . .      | 5 |
| MicrosoftWord . . . . .       | 6 |
| Known Problems . . . . .      | 6 |
| Useful Sites . . . . .        | 6 |

## Directory Structure

### Personal Directory

On the personal drive, there should be a file for ACIR documents. This is where you keep all the in progress documents. This is a place for anything that is related to the ACIR project that you are currently working on. There may be folders set up inside the folder for each type of document that you are working on. No one has access to these files except you.

### Shared Directory

*H://Dept/Govdocs/ACIR* is where completed documents are kept. Place the document here when it is completed and inform Cathy that it is done. She will then place it on the server from this location. Drive *G://* is a backup of *H://*.

## PDF Files

### File Size

File size is very important. In a PDF file that is placed on the Internet, the file size determines how long the user has to wait to download the document. The first page is especially important. If it does not load quickly, we may lose the user. In documents that have been scanned by an outside source, we have no say in the file size. In the documents that you scan, there is a process that will drastically reduce the file size, while keeping the quality of the document within reasonable limits.

### Byte Serving

Byte Serving is a feature from Adobe that is supported by our server. (Appendix A Section 1) It essentially means that, on the Internet, when a patron requests a document the server will push that document to the user a page at a time. This allows large files with many pages to load quickly because only one page loads at a time. The rest of the document does not load while the patron looks at any particular page. Links and bookmarks will work while byte serving. In order to implement byte serving, the user must have options on their Adobe plug-in configured correctly. This is explained in the supporting material on our site at: <http://www.library.unt.edu/gpo/acir/technicaldoc.htm> (Appendix A, Section 2).

In order to make it possible for the server to byte serve a document to a user, the document must be optimized (Appendix A, Section 3). In order to accomplish this, when saving a document use the Save As function. In the pop up box, make sure that Optimize is checked. Then select the location to place the document and click on save. This may take a couple of minutes. When it is done saving, the document is ready for byte serving.

The server automatically carries out this function. Some example sites that implement byte serving are included in (Appendix A, Section 4).

## Document Processing

There are two processes for creating effective online PDF files.

### Process 1

If the document is scanned by someone else and comes to us as a complete, bound, document then follow process 1.

1. Go to the download site: <http://plutonium-erl.actx.edu/contentdev/northtexas.html>
  2. Look for new documents that have not yet been downloaded.
  3. Download any of these documents and save them to your personal directory .
  4. Open Adobe Exchange.
  5. Open the file that you will work on in Adobe Exchange.
  6. Click the Fit Page button (middle page button) or select the View menu and select Fit Page.
  7. Click the Bookmarks with Page view (middle display button) or select the View menu and select Bookmarks and Page (Ctrl + 7).
  8.
    - A. Cycle through each page, bookmarking important pages (Ctrl + B) taking cues from the contents and naming the bookmarks accordingly.
    - B. As you cycle through the pages make sure that all the pages are there and crop any pages that do not contain a lot of outside white space. Cropping pages can be done by pulling down the Document menu and selecting Crop Pages. You may crop more than one page at a time, but it is usually more effective to crop one page.
- \* Note: Cropping does not delete the cropped part of the page, it only hides it to make it more attractive to the user. It does not affect download time.*
9.
    - A. Once all important pages are bookmarked, go to the Contents page and organize your bookmarks according to the order found there. You order bookmarks by clicking on the small page icon next to the bookmark to highlight it. Hold the mouse button down to grab it and drag it to where you want to place it.
    - B. Nest the bookmarks for items contained in the Contents under the Contents bookmark. To nest a bookmark use the same process that you used for ordering bookmarks, but instead of placing it on the same level as the bookmark above it, drag the selected bookmark to the right until the guide line is indented. Then let go of the mouse button. If you correctly nested the bookmark, an arrow will appear next to the upper level bookmark.



- C. To select several bookmarks, hold down the shift key while clicking on the page icons next to the bookmarks. You may then drag and drop many book marks at once.
10. Once the document has been bookmarked and you are on the Contents page, it is time to link the Contents.
    - A. Select the Link tool from the selection bar. It is the two chain links together .
    - B. Draw a box in the contents around the area of one of the sections. A Box will pop up.
    - C. On the first link, when the box pops up hit the "I" key. This will select an invisible rectangle from the type menu in the box. When linking to an area within the document you are working on, make sure Go to View is selected in the Type box. There are many other selections possible that provide for other functions, such as linking to another file, or for linking to the web. Make sure the magnification is set to Fit to Page. Then, using the contents or the page cycle keys, go to the page that you are linking to. When you find the place you want to link to, click the Set Link button. Continue this process for the entire Contents.
  11. If the document has an Index, then follow the above process and link the individual Index items.
  12.
    - A. Once the document is linked, go to the File menu and select Document Info
    - B. Choose Open under Document Info. This is where you determine how the document will look to the user when it is opened.
    - C. For our purposes select Bookmarks and Page for the Initial View.
    - D. Under Page Number select 1.
    - E. For Magnification select Fit Page.
    - F. For Page Layout stay with the default. Then click OK.
  13. Often the first few pages have stamps on them or other damaged areas. To correct this there needs to be some touch-up.
    - A. Select the form tool.
    - B. Draw a box around the damaged area. A menu box will pop up.
    - C. Name the area something simple ex: "w1" for white button 1.
    - D. Match the background color to the color of the selected area. If the area has many colors, you may need to draw many boxes to properly cover any damaged area.
    - E. Sometimes the letters of the title are partially covered. You can repair this by either inserting text, trying to match the font and color as closely as possible, or drawing many boxes and filling them with the text color to make it look like the letter. Examples of both text repair and of cover touch-up are on IPSFv18n3. Open this file and press the form button. The boxes will appear. Double click these boxes to look at the properties. It is often hard to match the color exactly on the cover. Use the custom colors, and use trial and error until the color matches.

14. Now go to the File menu, select Document Info, and choose General (Ctrl + D). This is where you set up how your document will appear to the search engines. Input the Title of the document, the Subject, the Author, and keywords that the search engine should bring your document up by (maximum 5000 characters), and click OK (Appendix C).
15. Save the document by selecting the Save As menu item and making sure that Optimize is checked. Then save it to the ACIR directory on the G drive.

*\*Note: Save often on your personal drive using Ctrl + S. This does not optimize the document, but it will ensure that you do not lose your work.*

### *Process 2*

If we only have the document in a paper, microfiche, or microfilm format then the process for placing it online is more lengthy, complicated, and time consuming.

1. If the document is on microform, a high-quality copy must be made for scanning.
2. Make sure that the paper document is as clean as it can be for the best possible scan.
3. You must be logged into the computer that is processing the scan so that you will have access to your network accounts.
4. For scanning we use two programs. We use PhotoShop and OmniPage Pro. OmniPage Pro is used for black and white documents. OmniPage is fast and simple to use.
  - A. Place the document on the scanner. If the document is oversized, try to fit the whole document on the scan surface, cutting off blank margins.
  - B. OmniPage defaults to scan at 300 dpi. Click on the Scan button.
  - C. Scan all the black and white pages of the document .
  - D. Go to File and select Save Image. Save them all as page images. Save them as TIFF files. Save them on your personal drive under a folder with the name of the document. When saving make sure that you have " Save All Pages" checked.
  - E. Do not OCR the pages in OmniPage. Name them BWtitle. OmniPage will number the pages starting with page 1. Even if you must skip pages OmniPage will still number consecutively. This is why you name them with the BW prefix.
5. PhotoShop is used for scanning pages with any background color that would cover up text, for any item with a table with background color, or for any item with color pictures (ex: the cover page). It is very important to keep the file size small, especially for the first page. In order to keep each color page small follow the following processes.

### Cover Page

- A. Go to the File menu and select Import, Twain 32.
- B. Set up the driver with the following settings. 72 DPI, Color Photo RGB.
- C. Press scan.
- D. Exit the Scan driver and crop the page to exclude ugly edges.
- E. Save as covertitle in personal directory in the same folder as the black and white scans.

### Other Pages

- A. Go to File, Import, Twain 32.
  - B. Set up the driver with the following settings. 200 DPI, B&W Photo.
  - C. Scan each page.
  - D. Exit the scan driver.
  - E. Crop the page.
  - F. Go to Image, Mode, Bitmap, 50% threshold and select OK.
  - G. Use the filter to make the image file size smaller.
  - H. Save as cpage# using the true page number of the scanned page in your personal directory as a TIFF file. This will help you later when binding.
6. Once all the images are saved on your personal directory, you can access them from the documents computer. They must be bound using Adobe Exchange.
    - A. Go to the File menu and select import image.
    - B. Insert the cover.
    - C. Continue inserting images in order, checking them against the print document to insure that they are in the proper order.
  7. Once the entire document is bound, save it as a PDF by the title in your personal directory.
  8. The next step is to make the text of the PDF searchable (Appendix B, Section 2)
    - A. Make sure that no other program is running on your computer.
    - B. Turn off your screen saver.
    - C. From the Document menu, select Capture Pages.
    - D. Go into Preferences in the pop up box and make sure that Original Text with Hidden Images is selected in the PDF Output Style box.
    - E. Because the cover is not at 200 DPI or greater, you will need to capture page through the end of the document.
    - F. You can stop capture and save it at the point that it stopped. You may then come back later and start from that point. Any page that has been captured cannot be captured again. Some pages cannot be captured due to errors on the page. There is nothing that can be done about this. Skip these pages, and continue to capture the rest of the document. The capture software that we are using is not very accurate. This prohibits exhaustive searching.

9. Once the document has been bound (step 5), and captured (step 7), it is in the same format as the downloaded documents from *Process 1*. Continue by starting with step 6 in *Process 1*.

## ADA Compliance

One of our goals is to make the documents that we are placing on the web accessible to the visually impaired. We have an adaptive computer lab on campus that currently runs the speech software "jaws." However, this software, and no other speech software, will read a PDF file, as it is an image with text in the background. In order to get the document in a format that the speech software can understand, follow this process (Appendix B, Section 1 ):

1. After a document has been through the capture process, under the Edit menu, select Copy File to Clipboard (Ctrl. + Shift + K).
2. Open a program that can read the RTF file format (ex: Word, WordPad, etc.)
3. Paste into this program (Ctrl. + V). The text from the document should appear on the screen.
4. Clean up any obviously bad areas.
5. Save this as ASCII text under the same name as the PDF on the shared directory .
6. Report the completed documents to Cathy.

## Microsoft Word

You can save a Word file as a PDF with the use of a plug-in. The plug-in introduces a box on the tool bar that will show a PDF symbol and give its status. To save a Word file to a PDF, simply click this button.

## Known Problems

### Unknown Error While Saving Cannot Save [9] or [10] (or similar error message)

This error usually occurs if you have made a lot of modifications to a document without saving using just the Ctrl. + S save. It will occur when trying to use the Save As feature during the final pass of Adobe's optimize cycle. After it happens you will not be able to save using either the Save or Save As. Fear not! Your document has been saved. It brings this error up

in the final stage of the optimize cycle where the program is trying to check for errors. Exit Exchange and reopen it. Open your document and then use the Save As menu item. It should work with no problems. There is no other way to recover the document.

## Page Contains Image Plus Text Cannot Capture

This is a strange error that seems to occur for no reason sometimes. Other times it occurs because you have already captured this page, or the page has images that have been placed on it. I have not found a good solution for this problem. According to my sources, the best thing to do seems to be to skip the page with the problem and continue capturing the next page .

## Document is not byteserving when tested on the Internet

There could be a couple of reasons for this. The first is server side. The document must be optimized. You can check to see if a document is optimized by selecting Document Info under the File menu and choosing General. At the bottom of the pop-up box it will say whether the document is optimized. If it is not, pull it off the shared directory and do a Save As with optimize checked. Another problem can be that the client does not have the proper configuration selected for byteserving. Instructions regarding this are at: <http://www.library.unt.edu/gpo/acir/technicaldoc.htm>.

## Useful Sites

**PDF Archive Mailing List Interface:** <http://www.pdfzone.com/cgi-bin/wilma.cgi/pdf>

This is a wonderful site! This is usually the first place to go for answers to your questions. It is searchable and archived from the beginning. It is a list for answering questions from developers of PDF archives like ours. They answer many questions regarding PDF's but most questions regarding software development questions are on another list. This is a good thing because there is less clutter on this list-serv.

**PDF ResearchCompanion:** <http://www.performancegraphics.com/>

This site has some problems with their frames. You will need to bust the frames using the right mouse button to Open in a New Window. The site contains a wealth of informaiton.

**Adobe:** <http://www.adobe.com/homepage.shtml>

This site is also a great resource. The site is searchable and this is usually my first stop for information regarding specific error messages. There are many technical and white papers. There are also error message reports, all of which are called up by the search of the entire Adobe site.

**PDFZone:** <http://www.pdfzone.com/>

This is also another good site that is especially useful for finding plug-ins that are **PDF** specific. The site has other uses as well, and is often a good source for answering some up-to-date questions regarding the latest releases of Adobe Exchange/Acrobat related software.

# **Appendix B**

## **Search Engines**

| <b>Software</b>                           | <b>Price</b>   | <b>Positives</b>   | <b>Negatives</b>  |
|---|--|--|---|
| <b>Infoseek:<br/>Ultraseek<br/>Server</b> | Price: \$995- \$4,995 minus<br>the GSA or Educational<br>Discounts | Supports MS Office and PDF files<br><br>Easy Administration<br>Automatic Reindexing<br>Proper Name Recognition<br>Search as a Phrase<br>Search with Boolean operators<br>Search with Natural Language<br>Processing<br>Y2K compliance<br>Control summary information with<br>metatags<br>Customizable<br>Automatically finds spacing variants in<br>search terms.<br>Multiple language searches. (English,<br>Dutch, French, German, Italian,<br>Portuguese, Spanish mainly)<br>Provides a relevancy ranking.<br>Query refinement (search only these<br>results) |   |
| <b>Microsoft<br/>Index Server</b>         | Price: Free  | It is free if we have a Win NT 4.0 server<br><br>Supports PDF and MS Office formats<br>Most closely integrated with Windows<br>NT<br>Automatic Reindexing<br><br>MS is continuing development of the<br>search technology. Any future releases<br>will probably also be free.<br>Can control summaries with the use of<br>metatags   | It is not as advanced as commercial<br>search engines.<br>Does not rank the search results.<br>No duplication detection.<br><br>Often included HTML characters<br>when creating summaries<br>No Natural Language Processing |



| <b>Software</b>        | <b>Price</b>   | <b>Positives</b>  | <b>Negatives</b>  |
|------------------------|--|---|---|
| <b>Verity Search97</b> | Price: For the core Information Server pricing begins at \$7,995. The actual cost for implementation on an entire site would be considerably more. | Supports PDF and MS Office formats as well as over 200 other file formats.<br><br>Provides special filters for searching non-text documents.<br><br>Other companion products extend its capability.<br>Can control summaries with the use of metatags.<br>Provides a relevancy ranking. | Does not have Automatic Reindexing. If anything is added or deleted, the entire site must be reindexed.<br><br>Search 97 is really a suite of products. To purchase everything that we would need would be considerably expensive, and we may end up with a lot of things that we don't need. |

| <b>Search Engines Not Chosen</b> | <b>Reason</b>                                |
|----------------------------------|--|
| <b>SWISH-E</b>                   | Unix Platform                                |
| <b>Excite for Web Servers</b>    | Does not search PDF files.                   |
| <b>Harvest</b>                   | Unix Platform<br>Project is bankrupt         |
| <b>Sage (NUD*IST 4)</b>          | Project based search for personal computers. |
| <b>Excalibur</b>                 | Extremely high pricing.                      |
| <b>RetrievalWare</b>             | May be more than we need.                    |

# **Appendix C**

## **Scanning Log**

# Electronic Resource Library

---

## Scanning record log: North Texas State University

| <u>Document title</u>                                   | <u>File name</u>    | <u>Tiff Rages</u> | <u>Color Rages</u> |
|---|---------------------|-------------------|--------------------|
| Intergovernmental Perspective, Vol. 13, no. 3/4         | t00001              | (Same as too021 ) |                    |
| Significant Features of Fiscal Federalism, Vol. 1: 1991 | t00002              | 188               | 0                  |
| Significant Features of Fiscal Federalism, Vol. 1: 1990 | t00003              | 160               | 4                  |
| Significant Features of Fiscal Federalism, Vol. 2: 1990 | t00004              | 280               | 4                  |
| Significant Features of Fiscal Federalism, Vol. 1: 1992 | t00005              | 192               | 4                  |
| Significant Features of Fiscal Federalism, Vol. 2: 1992 | t00006              | 368               | 4                  |
| Significant Features of Fiscal Federalism, Vol. 2: 1991 | t00007              | 340               | 4                  |
| Significant Features of Fiscal Federalism, Vol. 1: 1995 | t00008              | 172               | 4                  |
| Significant Features of Fiscal Federalism, Vol. 2: 1994 | t00009              | 284               | 0                  |
| Significant Features of Fiscal Federalism, Vol. 1: 1994 | t00010              | 180               | 0                  |
| Intergovernmental Perspective, Vol. 10, no. 1           | t00011              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 10, no. 2           | t00012              | 35                | 1                  |
| Intergovernmental Perspective, Vol. 10, no. 3           | t00013              | 31                | 1                  |
| Intergovernmental Perspective, Vol. 11, no. 1           | t00014              | 47                | 1                  |
| Intergovernmental Perspective, Vol. 11, no. 2/3         | t00015              | 23                | 1                  |
| Intergovernmental Perspective, Vol. 11, no. 4           | t00016              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 12, no. 1/2         | t00017              | 35                | 1                  |
| Intergovernmental Perspective, Vol. 12, no. 3           | t00018              | 31                | 1                  |
| Intergovernmental Perspective, Vol. 13, no. 1           | t00019              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 13, no. 2           | t00020              | 31                | 1                  |
| Intergovernmental Perspective, Vol. 13, no. 3/4         | t00021              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 14, no. 1           | t00022              | 27                | 1                  |
| Intergovernmental Perspective, Vol. 14, no. 2           | t00023              | 23                | 1                  |
| Intergovernmental Perspective, Vol. 14, no. 3           | t00024              | 27                | 1                  |
| Intergovernmental Perspective, Vol. 14, no. 4           | t00025              | 27                | 1                  |
| Intergovernmental Perspective, Vol. 15, no. 1           | t00026              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 15, no. 2           | t00027              | 35                | 1                  |
| Intergovernmental Perspective, Vol. 15, no. 3           | t00028              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 15, no. 4           | t00029              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 16, no. 1           | t00030              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 16, no. 2           | t00031              | 27                | 1                  |
| Intergovernmental Perspective, Vol. 16, no. 3           | t00032              | 31                | 1                  |
| Intergovernmental Perspective, Vol. 16, no. 4           | t00033              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 17, no. 1           | t00034              | 51                | 1                  |
| Intergovernmental Perspective, Vol. 17, no. 2           | t00035              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 17, no. 3           | t00036              | 47                | 1                  |
| Intergovernmental Perspective, Vol. 17, no. 4           | t00037              | 59                | 1                  |
| Intergovernmental Perspective, Vol. 18, no. 1           | t00038              | 47                | 1                  |
| Intergovernmental Perspective, Vol. 18, no. 2           | t00039              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 18, no. 3           | t00040              | 31                | 1                  |
| Intergovernmental Perspective, Vol. 18, no. 4           | t00041              | 35                | 1                  |
| Intergovernmental Perspective, Vol. 19, no. 1           | t00042              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 19, no. 2           | t00043              | 43                | 1                  |
| Intergovernmental Perspective, Vol. 19, no. 3           | t00044              | 35                | 1                  |
| Intergovernmental Perspective, Vol. 20, no. 1           | t00045              | 47                | 1                  |
| Intergovernmental Perspective, Vol. 20, no. 2           | t00046              | 39                | 1                  |
| Intergovernmental Perspective, Vol. 20, no. 3           | t00047              | 43                | 1                  |
| <b>Totals:</b>  | <b>45 documents</b> | <b>3539</b>       | <b>61</b>          |

# **Appendix D**

## **ACIR Project Data**

## ACIR Project Data

### Outsourced Scanning of Intergovernmental Perspectives

| Title        | Downloaded | Bookmark and Page Setup | Contents Linked | Cover Touch-Up | Saving | Total | File Size (M) | Min/Sec | Pages | Min/Sec |
|--------------|------------|-------------------------|-----------------|----------------|--------|-------|---------------|---------|-------|---------|
| IPSPv10r184  | 17         | 6                       | 3               | 2              | 2      | 30    | 3.52          | 6:52    | 45    | 0:75    |
| IPSPv10r284  | 19         | 7                       | 3               | 1              | 2      | 31    | 2.81          | 11:03   | 38    | 0:55    |
| IPSPv10r384  | 16         | 8                       | 5               | 1              | 2      | 32    | 2.75          | 11:58   | 32    | 1:00    |
| IPSPv11r180  | 14         | 4                       | 5               | 2              | 2      | 28    | 4.40          | 6:36    | 48    | 0:58    |
| IPSPv11r2080 | 10         | 5                       | 5               | 1              | 2      | 32    | 3.53          | 10:56   | 24    | 1:33    |
| IPSPv11r480  | 21         | 6                       | 7               | 2              | 2      | 38    | 3.55          | 10:72   | 40    | 0:55    |
| IPSPv14r180  | 8          | 3                       | 2               | x              | 2      | 15    | 2.70          | 5:58    | 38    | 0:54    |
| IPSPv14r280  | 5          | 4                       | 3               | 3              | 2      | 17    | 1.12          | 8:52    | 24    | 0:71    |
| IPSPv14r380  | 7          | 5                       | 3               | 2              | 2      | 19    | 2.25          | 8:59    | 38    | 0:55    |
| IPSPv14r480  | 5          | 4                       | 3               | 2              | 2      | 17    | 2.44          | 8:37    | 28    | 0:51    |
| IPSPv15r180  | 8          | 3                       | 2               | 2              | 2      | 18    | 3.49          | 5:17    | 40    | 0:45    |
| IPSPv15r280  | 13         | 4                       | 2               | 2              | 2      | 23    | 3.54          | 7:57    | 36    | 0:54    |
| IPSPv15r3    | 10         | 4                       | 2               | 2              | 2      | 20    | 3.54          | 5:49    | 40    | 0:50    |
| IPSPv15r4    | 12         | 3                       | 5               | 1              | 4      | 27    | 3.80          | 7:11    | 40    | 0:55    |
| IPSPv15r1    | 10         | 6                       | 5               | x              | 2      | 23    | 3.25          | 7:21    | 40    | 0:55    |
| IPSPv15r2    | 10         | 4                       | 5               | x              | 2      | 21    | 4.07          | 5:15    | 48    | 0:44    |
| IPSPv15r3    | 15         | 6                       | 6               | x              | 2      | 29    | 3.87          | 7:49    | 32    | 0:51    |
| IPSPv15r4    | 20         | 7                       | 6               | x              | 2      | 35    | 3.36          | 10:42   | 40    | 0:55    |
| IPSPv17r1    | 5          | 6                       | 5               | x              | 2      | 21    | 4.34          | 4:54    | 52    | 0:40    |
| IPSPv17r2    | 4          | 8                       | 7               | x              | 2      | 23    | 3.18          | 7:20    | 40    | 0:55    |
| IPSPv17r3    | 6          | 8                       | 7               | x              | 2      | 25    | 3.95          | 6:33    | 48    | 0:52    |
| IPSPv17r4    | 20         | 8                       | 8               | 2              | 2      | 38    | 4.98          | 7:53    | 60    | 0:53    |
| IPSPv18r1    | 20         | 9                       | 4               | 10             | 2      | 42    | 4.06          | 10:34   | 48    | 0:58    |
| IPSPv18r2    | 20         | 5                       | 4               | 2              | 2      | 34    | 3.40          | 10:30   | 40    | 0:55    |
| IPSPv18r3    | 20         | 5                       | 7               | 8              | 2      | 41    | 2.80          | 14:44   | 32    | 1:28    |
| IPSPv18r4    | 22         | 4                       | 3               | 8              | 2      | 39    | 2.95          | 13:54   | 36    | 1:08    |
| IPSPv18r1    | 15         | 3                       | 4               | 8              | 2      | 35    | 3.97          | 9:55    | 40    | 0:58    |
| IPSPv18r2    | 12         | 5                       | 8               | 8              | 2      | 36    | 3.75          | 9:73    | 44    | 0:52    |
| IPSPv18r3    | 17         | 5                       | 8               | x              | 2      | 32    | 3.35          | 9:75    | 36    | 0:58    |
| IPSPv20r1    | 10         | 4                       | 2               | 1              | 2      | 27    | 4.58          | 6:18    | 48    | 0:56    |
| IPSPv20r2    | 23         | 4                       | 3               | 8              | 2      | 39    | 3.08          | 12:46   | 40    | 0:55    |
| IPSPv20r3    | 21         | 2                       | 2               | 3              | 2      | 32    | 5.14          | 9:55    | 44    | 0:58    |
| IPSPv12r1e2  | 15         | 4                       | 3               | 1              | 2      | 25    | 5.03          | 6:23    | 36    | 0:59    |
| IPSPv12r2    | 12         | 4                       | 2               | 2              | 2      | 23    | 2.82          | 7:55    | 32    | 0:59    |
| IPSPv13r1    | 12         | 3                       | 2               | 2              | 2      | 21    | 5.07          | 6:54    | 48    | 0:53    |
| IPSPv13r2    | 14         | 4                       | 3               | 4              | 2      | 27    | 6.18          | 4:38    | 32    | 0:54    |
| IPSPv13r3e4  | 16         | 4                       | 2               | 2              | 2      | 26    | 5.35          | 6:75    | 48    | 0:70    |
| Total        | 530        | 186                     | 190             | 21             | 75     | 1038  | 127.24        |         | 1432  |         |
| AVG          | 14.32      | 5.30                    | 4.95            | 2.35           | 2.05   | 28.05 | 3.44          | 8.40    | 38.70 | 0.74    |

Fig. 1

## ACIR Project Data

### inning of Intergovernmental Perspectives

| Scanning | OCR   | Bookmark and Page Setup | Contents Linked | Cover Touch-Up | Saving | Total  | File Size (M) | Min/M | Pages | Min/Pg |
|----------|-------|-------------------------|-----------------|----------------|--------|--------|---------------|-------|-------|--------|
| 75       | 40    | 15                      | 2               | 1              | 4      | 137    | 5.01          | 27.95 | 32    | 4.28   |
| 95       | 45    | 9                       | 3               | 3              | 3      | 158    | 13.20         | 11.97 | 48    | 3.29   |
| 40       | 45    | 8                       | 3               | 2              | 2      | 100    | 11.10         | 9.01  | 36    | 2.78   |
| 25       | 25    | 8                       | 2               | 1              | 2      | 63     | 6.63          | 9.50  | 32    | 1.97   |
| 72       | 35    | 4                       | 2               | 1              | 2      | 116    | 3.51          | 33.05 | 32    | 3.63   |
| 55       | 62    | 4                       | 2               | 1              | 2      | 126    | 4.64          | 27.16 | 44    | 2.86   |
| 60       | 72    | 6                       | 4               | 2              | 2      | 146    | 3.69          | 39.57 | 36    | 4.06   |
| 72       | 86    | 4                       | 3               | 1              | 2      | 168    | 3.01          | 55.81 | 32    | 5.25   |
| 30       | 30    | 4                       | 2               | 1              | 2      | 69     | 4.75          | 14.53 | 32    | 2.16   |
| 40       | 50    | 4                       | 3               | 4              | 3      | 64     | 3.24          | 19.75 | 32    | 2.00   |
| 564      | 490   | 66                      | 26              | 17             | 24     | 1147   | 58.78         |       | 366   |        |
| 58.40    | 49.00 | 6.60                    | 2.60            | 1.70           | 2.40   | 114.70 | 5.88          | 24.77 | 35.60 | 3.23   |

Fig. 2

## ACIR Project Data

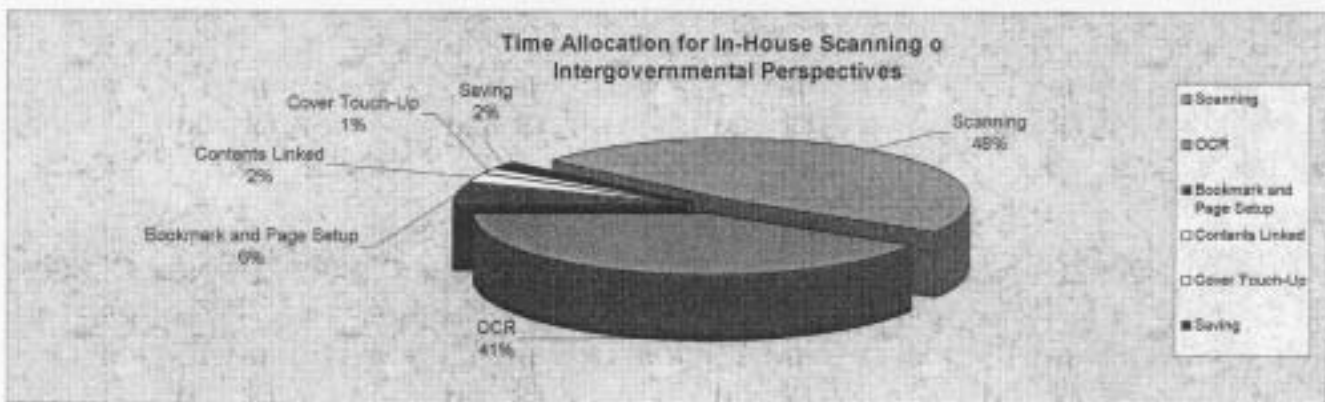
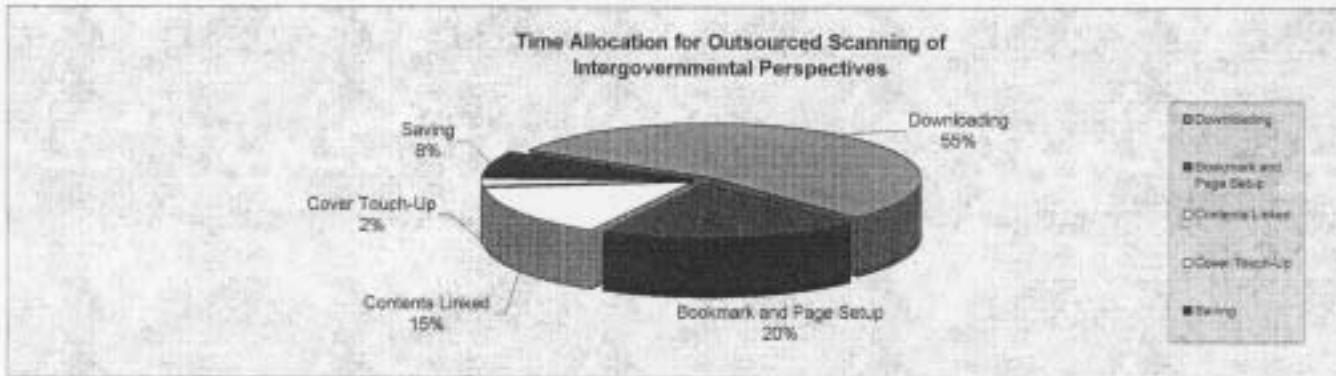


Fig. 3

# ACIR Project Data

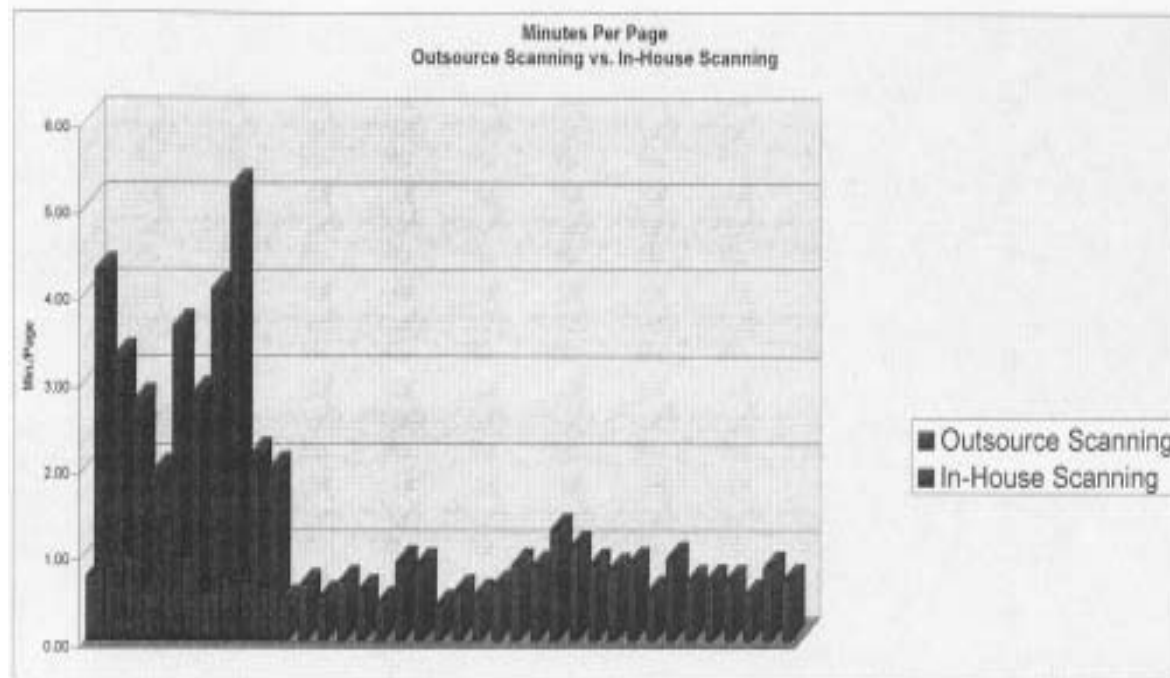


Fig. 4

Data from Fig. 1 and Fig. 2.



### ACIR Project Data

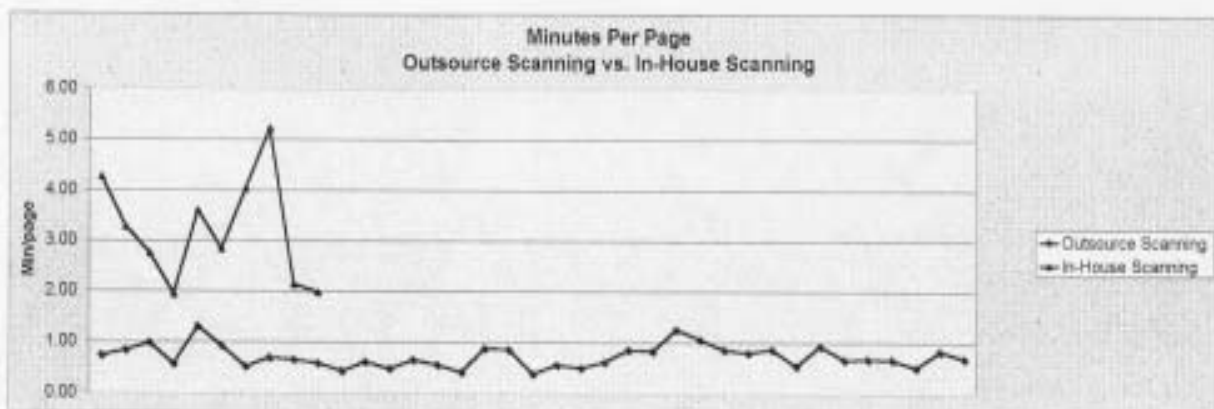


Fig. 5

\*Data from Fig. 1 and Fig. 2.

### ACIR Project Data

#### Outsourced Scanning of SFFF

| Title            | Download Time | Booklist and Page Setup | Contents Linked | Index Linked | Saving      | Total         | File Size (M) | Min/M        | Pg        |
|------------------|---------------|-------------------------|-----------------|--------------|-------------|---------------|---------------|--------------|-----------|
| SFFF 1991 Vol. 1 | 15            | 90                      | 45              | 135          | 5           | 290           | 12.04         | 28.80        | 1         |
| SFFF 1990 Vol. 1 | 19            | 112                     | 28              | 30           | 5           | 191           | 7.84          | 25.00        | 1         |
| SFFF 1990 Vol. 2 | 25            | 35                      | 85              | 45           | 5           | 175           | 13.58         | 13.09        | 2         |
| SFFF 1992 Vol. 1 | 14            | 100                     | 30              | 45           | 5           | 194           | 9.46          | 20.51        | 1         |
| SFFF 1992 Vol. 2 | 20            | 105                     | 55              | 190          | 5           | 385           | 16.67         | 22.82        | 3         |
| SFFF 1991 Vol. 2 | 23            | 120                     | 30              | 115          | 5           | 293           | 18.25         | 18.03        | 3         |
| SFFF 1994 Vol. 1 | 12            | 25                      | 10              | 26           | 5           | 78            | 6.23          | 9.48         | 1         |
| SFFF 1994 Vol. 2 | 25            | 30                      | 43              | 78           | 5           | 181           | 12.25         | 14.78        | 2         |
| SFFF 1995 Vol. 1 | 18            | 20                      | 8               | 19           | 5           | 69            | 6.03          | 8.47         | 1         |
| <b>Total</b>     | <b>179</b>    | <b>637</b>              | <b>311</b>      | <b>683</b>   | <b>40</b>   | <b>1905</b>   | <b>102.15</b> |              | <b>21</b> |
| <b>Avg</b>       | <b>19.89</b>  | <b>70.78</b>            | <b>34.59</b>    | <b>75.99</b> | <b>5.00</b> | <b>206.11</b> | <b>11.35</b>  | <b>18.16</b> | <b>24</b> |

Fig. 6

### ACIR Project Data

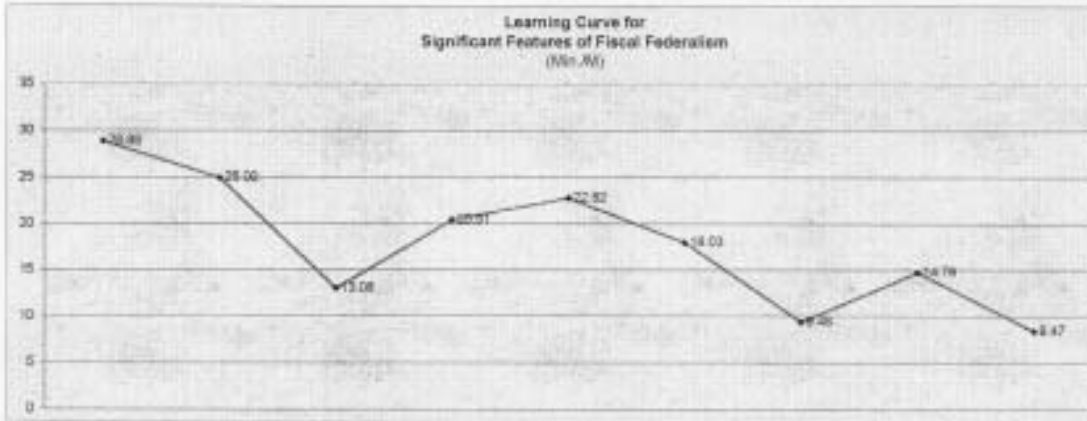


Fig. 7

# ACIR Project Data

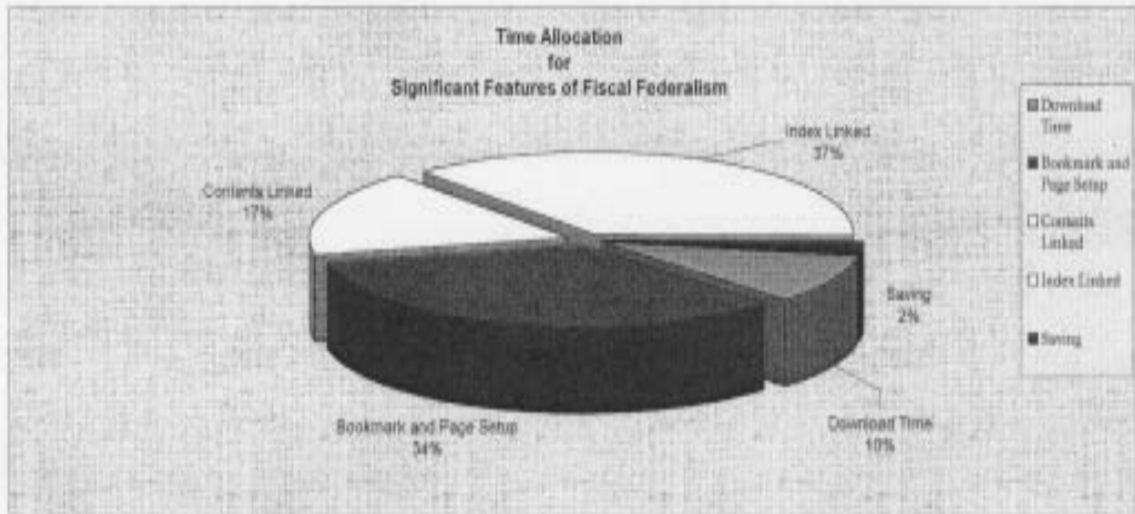


Fig. 8

Data from Fig. 8