

Accounting Historians Journal

Volume 33
Issue 2 December 2006

Article 8

2006

Digitizing the Accounting Historians Journal: A short history

Royce D. Kurtz

David K. Herrera

Stephanie D. Moussalli

Follow this and additional works at: https://egrove.olemiss.edu/aah_journal

Part of the [Accounting Commons](#), and the [Taxation Commons](#)

Recommended Citation

Kurtz, Royce D.; Herrera, David K.; and Moussalli, Stephanie D. (2006) "Digitizing the Accounting Historians Journal: A short history," *Accounting Historians Journal*: Vol. 33 : Iss. 2 , Article 8.

Available at: https://egrove.olemiss.edu/aah_journal/vol33/iss2/8

This Article is brought to you for free and open access by the Archival Digital Accounting Collection at eGrove. It has been accepted for inclusion in Accounting Historians Journal by an authorized editor of eGrove. For more information, please contact egrove@olemiss.edu.

Accounting Historians Journal
Vol. 33, No. 2
December 2006
pp. 157-168

Royce D. Kurtz
UNIVERSITY OF MISSISSIPPI
David K. Herrera
UNIVERSITY OF MISSISSIPPI
and
Stephanie D. Moussalli
UNIVERSITY OF WEST FLORIDA

DIGITIZING THE ACCOUNTING HISTORIANS JOURNAL: A SHORT HISTORY

Abstract: The University of Mississippi Library has digitized the *Accounting Historians Journal* from 1974 through 1992, cover-to-cover. The American Institute of Certified Public Accountants' gift of their library to the University of Mississippi was, fortuitously, the impetus for the *AHJ* digitizing project. A complicated chain of events followed which included discussions with the Academy of Accounting Historians for copyright permission, an application for a federal grant, negotiations with software vendors, and decisions about search capabilities and display formats. Each article in *AHJ* is now full-text searchable with accompanying PDF page images.

The story of how the University of Mississippi Library (UML) came to digitize the early years of the *Accounting Historians Journal* (*AHJ*) begins with a fortuitous series of events involving four organizations¹:

- the American Institute of Certified Public Accountants (AICPA), a nonprofit, professional organization
- the Institute of Museums and Library Services (IMLS), a federal government library and museum service organization

¹The abbreviations and acronyms used in this article are listed and defined in Appendix A.

Acknowledgments: The authors gratefully acknowledge the comments of two anonymous referees and the participants at the October 2005 Academy of Accounting Historians Annual Conference, where an earlier version of the paper was presented. We also thank the Institute of Museum and Library Services for the grant that made the digitization possible.

- Innovative Interfaces, Inc. (III), a private library software company
- the Academy of Accounting Historians (AAH), a non-profit, academic organization that publishes *AHJ*

Over a five-year period, the disparate interests of these organizations converged at the UML and produced, in September 2005, open, public access to the first 19 volumes (1974-1992) of the digital *AHJ*.² For the UML, *AHJ* was the first product in its new Digital Accounting Collection (DAC). Digitally publishing the journal was an intensive introduction to the scarcely charted seas of research collections on the Internet.

THE UML ACQUIRES THE AICPA'S LIBRARY

In September 2000, James Davis and Dale Flesher, respectively the dean and associate dean of the University of Mississippi School of Accountancy, approached John Meador, the dean of university libraries, with an audacious proposition. They proposed that the university enter the competition for the enormous accounting library of the AICPA. The AICPA had decided to divest itself of its venerable library, a collection begun in 1918 [Neloms, 1987; Anonymous, 1998]. According to the Institute's *CPA Letter*, the AICPA was moving "toward providing more services in an electronic environment. Part of the evolution from a paper-based environment to an electronic one includes the relocation of the AICPA library collection" [Anonymous, 2001]. A new home at a major university was needed. A competitive process was established and requests for proposals were sent to major universities with distinguished accountancy programs.

When Dean Davis received an invitation to bid from the AICPA [Rothberg, 2000], he saw a wonderful opportunity to add tens of thousands of items to the university's already large accounting collection. At the library, John Meador was enthusiastic about the prospect of adding a world-class collection to the library's holdings. The two deans approached the University of Mississippi Foundation and Robert Khayat, chancellor of the university. The reception was enthusiastic and the university's resources were pledged to the proposal. The Robert M. Hearin Foundation was persuaded to underwrite some of the expenses of housing and re-cataloguing the collection. A proposal was drafted, "Plan for a National Library of the Accounting Profession: A Proposal to the American Institute of Certified Public

²available at <http://umiss.lib.olemiss.edu:82/screens/dacopac.html>

Accountants” [University of Mississippi, 2000]. The document included plans for maintaining and preserving the collection and for providing the AICPA’s membership with continued access and research services. Also mentioned were plans for the digitization of rare items. Chancellor Khayat flew to New York with Dean Meador to present the proposal to the executive board of the AICPA. Representatives of the AICPA Foundation made site visits to each of the competing institutions.³

In February 2001, the University of Mississippi’s proposal won the award. A collection of 33,000 books, 93,000 pamphlets, 1,300 periodical titles, over 500 photographs, and 191 rare books – the “National Library of the Accounting Profession” – was now the property of the University of Mississippi [Anonymous, 1998; Rothberg, 2000]. When merged with the library’s previous accounting materials, the largest accounting collection in the world had been amassed.

On June 27, 2001, the final contract was signed. Just over a month later, the collection arrived in Oxford, Mississippi. A dozen semi-trucks carrying over 4,000 boxes of material rolled into a temporary warehouse on the outskirts of Oxford. Library staff now had to unpack, process, and integrate the materials into the library holdings. The rare books and photographs were immediately housed with the library’s special collections. New shelving was erected and large sections of the library’s materials were shifted to make room for the new arrivals. Rapid cataloguing of the most frequently used items began. Simultaneously, library staff started providing research service to the AICPA’s membership by answering reference questions and loaning books and articles from the collection (Martin, 2004).

THE IMLS GRANT TO THE LIBRARY

While no formal agreement had been reached between the AICPA and the university in regards to digitizing any of the collection, both parties had discussed the possibility, and the university had included a plan for digitizing a portion of the material in its original proposal [University of Mississippi, 2000]. Soon after winning the award of the AICPA’s collection, the university decided to pursue federal funding for a digitization program. In January 2002, with the assistance of Dean Davis, the UML secured a \$350,000 directed grant from the federal government’s

³Other competing libraries included Baruch College (CUNY), the University of Illinois, Michigan State University, and Ohio State University.

Institute of Museums and Library Services ["An Act," 2212].⁴ For the IMLS, this was an opportunity to encourage an academic library to provide wider public access to its research materials through the World Wide Web. The grant was to be exploratory, designed to provide an opportunity for the library to set up equipment and procedures to start a digitization program.

THE UML'S PARTNERSHIP WITH III

Now that the library had the funding to begin a digitization project, it had to select software to manage the digitized material. Such software handles the housing of the collection, provides public access, maintains the bibliographic records, and provides tools for searching the material. Ideally, the software also integrates smoothly with the library's existing information technology platform.

University libraries often turn to private companies for their integrated library system needs, ranging from basic online catalogue services to federated database searching and circulation management. As libraries' services and goals evolve, companies routinely develop new products, including, in recent years, digital collection management software. The UML's integrated library system platform is provided by III. In 2002, when library staffers were exploring digital library products, they learned that III planned to create a digital collection management suite. The suite was to be built in part upon technology already in use at the library, in particular the online public access catalogue (OPAC). However, part of the suite was still just etherware – it had yet to be created. This was the "Metadata Builder" (MB) core of the product. Other components included Millennium Media Management and a separate database dedicated to the digital collection.

III needed development partners, a small group of university libraries willing to create their digital collections at the same time as the software was being designed. III offered a good price to potential partners, and the University of Mississippi Library accepted the offer. Thus, part of the IMLS grant funded the digital management suite; another piece of the funding went to purchase an additional database dedicated to the new collection.

⁴For readers consulting the statute online, the grant to the University of Mississippi appears in the statute under Title IV: Related Agencies; Institute of Museum and Library Services: Grants and Administration, at 115 STAT2212.

There were advantages and disadvantages to choosing the III product. It had the great advantage of being easily integrated with other catalogue modules the library already uses. Much of the search software and its functionality was copied straight from the online catalogue and so had already proven its worth. Furthermore, joining a software project at its earliest stage allowed the library to influence many aspects of how the product would finally perform. On the other hand, a development partner must spend a great deal of time testing and debugging as the module moves from its raw, rudimentary stages to a polished final product. Indeed, the testing and the reporting of bugs was a continuous part of the process from August 2004 through September 2005.

A PROPOSAL FROM THE AAH TO THE UML

Having acquired a large collection, a sum of money for digitizing, and a software development partner, the next question was what to digitize. The editors of *AHJ* had heard about the library's nascent digitizing project and, in the summer of 2002, an informal group consisting of Bill Samson, Gary Previts, and Dale Flesher approached the library about digitizing the early issues of *AHJ*. For the AAH, this was an opportunity to disseminate early volumes of its flagship journal much more widely. Volumes after 1992 had already been digitally published by commercial firms, but the years from 1974 to 1992 existed only in hard copy. As for the UML, *AHJ* was a cornerstone in its vision for an online historical accounting collection (later baptized the Digital Accounting Collection). The journal offered practical advantages as well. For an initial foray into digitizing, it is preferable to use fairly uniform, easily read material. The individual articles in a journal, far shorter than monographs, allow an easy manipulation of files. Spare copies of the journal were made available for the project, obviating the risk of destroying irreplaceable material while digitizing it. A November 25, 2002 letter from the then AAH president, Bill Samson, provided formal permission for the UML to proceed [Samson, 2002].

PHASE 1: DIGITIZING *AHJ*

In many senses the library had started backwards. It had received a grant to digitize before it had actually outlined a program, chosen materials, or identified any procedures. But now that the Academy had granted permission to proceed with *AHJ*,

the library could select scanning hardware and software and hire staff.

The hardware was quickly purchased – a computer with an enormous hard drive (one hundred gigabytes) and a simple scanner. Large, expensive scanners are available that handle fragile archival material and scan high volumes of material automatically, but the library's initial digitization project did not require such equipment. Indeed, at the earliest stages, no dedicated scanner was used at all. Instead, an existing printer-scanner was used from which files were faxed to the project computer for processing. However, the multiple steps involved proved cumbersome and inefficient, so a separate scanner was finally purchased.

Software required more thought. It would have to perform optical character recognition (OCR), create professional-quality PDF files, including some graphics, and make simple text files. Several software products were considered and tested, with AbbyFine Reader finally selected for the project. AbbyFine proved better than competing software at a number of tasks. It recognized bold and italic text more often, handled varied font sizes better, and more frequently ignored stray marks and flaws on the document pages. It also had a much lower error rate in making the text files; it recombined hyphenated words so they could be found in searches, and it maintained text integrity when images had to be eliminated, so that text was seldom scrambled or lost. For these reasons, AbbyFine Reader required far less manual editing than did competing products. Student workers also found Abby relatively easy to set up and learn.

In 2003, two students were hired, an undergraduate to scan pages and a graduate student in computer science to digitize volumes 1-19 of *AHJ* with AbbyFine Reader. Faculty assisted in these efforts. The PDF files were then checked for errors. Over 3,500 pages were involved in this phase of the project.

The process was a serious hands-on experience both in workflow management and in determining accurately the capabilities of OCR software. For each article in *AHJ*, at least five files were created: a searchable PDF file, a text file, an image file, and two files containing images of the relevant issue cover. The PDF file would be the main document used by the public. The underlying text file, inaccessible to the public, is what the catalogue search engine would use when commanded to search for a particular name or word. For scholarly accuracy, an unedited file containing simple images of the pages in the article would

also be available to the public. Scholars might then immediately check the “original” (i.e., unprocessed) document when questions arose concerning the accuracy of the searchable PDF file. Finally, JPEG files containing images of the issue covers were made to enhance the bibliographic records that were later created.

The ongoing checking of text for errors over the course of a year and a half revealed several points. First, the OCR program was remarkably accurate. Error rates were extremely low, a mistake being located in every three to five pages of text. Mistakes were both substantive (errors in letter and numeral interpretation) and stylistic (failure to identify correctly type size or font). Occasionally, the machine omitted an entire line. Human errors also occurred, including poorly scanned and missing pages, and, rarely, missing articles. Second, some errors were missed even with systematic checking. This was revealed when a second quality control check was performed on most articles. Third, the librarians learned first hand the complexity of a published “text.” *AHJ* was certainly more complicated than anyone had originally thought. Even articles containing nothing but text had a variety of font sizes and styles, foreign words with accents and non-Roman scripts, and special characters and symbols.

Finally, tables, figures, and photographs were more common and problematic than originally expected. Tables generally required extensive manual editing. Figures sometimes could not be read by the OCR program at all and had to be inserted as images. Furthermore, software products that accurately perform OCR often create very mediocre reproductions of images, particularly photographs. Fortunately, the earliest years of *AHJ* had few photographic reproductions, but in later issues, image quality became a serious quality control issue. While AbbyFine Reader’s OCR program works better in black and white, images reproduce best in grayscale. Adjusting for this improved many of the photographs, but clean-up with Adobe Photoshop was still required for others. These images were then reinserted into the digital copies. Whether processed with Photoshop or with Abby, the images were reproduced separately in the end and inserted back into the PDF in the final stage of processing.

The production of a searchable PDF copy of each article in *AHJ* was completed in August 2004. Error-checking, final clean-up, and creating the associated text, image, and JPEG files continued for another three to four months.

PHASE II: CREATING PUBLIC ACCESS TO THE DIGITAL *AHJ*

While one group of librarians was worrying about creating digital images, another group was working on issues of metadata and digital collection management with III in its development of the MB software. In August and September 2004, several preliminary steps were completed. A new database was profiled on the library's server and several new III modules were installed: Advanced Keyword Indexing, Millennium Media Management, and MB. Librarians worked closely with the development team at III from August 2004 through September 2005 to complete the development phase and the beta test phase for MB, the core digital collections manager product.

MB treats standard library catalogue entries as a form of metadata in that they describe the base item, be it a book, an article, a film, or a photograph. MB would allow the librarians to catalogue the DAC by creating a separate bibliographic record for each article or other item (such as tables of contents) in *AHJ*. The Millennium Media Management module would then store all the files for each article, linked to the relevant bibliographic record. The OPAC function would allow users to search the catalogue records, and the advanced keyword indexing function would allow users to search the full text of the documents as well.

The first step for the library in working with MB was the selection of a metadata language in which to make the bibliographic entries. The larger library community had realized that the traditional rules for cataloguing books⁵ were not necessarily the best rules for cataloguing the rapidly expanding digital world, where the full text was often immediately available. A different, briefer bibliographic standard for digital material (i.e., a metadata description) was needed. The Dublin Core (DC) standard, later revised and expanded into "qualified DC," was created in 1995 for this purpose. The UML and III selected qualified DC as the metadata standard for cataloguing the DAC. The qualified DC has about fifty fields and subfields, such as title, creator, publisher, etc. A formal library committee went to work selecting which of the fifty-odd fields to use and devising standard wording for fields pertaining specifically to digital characteristics, such as dots per inch and file size in kilobytes

⁵The traditional standard follows the AARC2 protocol for creating MARC records.

or megabytes. The committee established general guidelines for inputting DC records, wrote a short in-house, how-to manual based on the Western States Dublin Core Metadata Best Practices guidelines [John Davis Williams Library, 2005], and created a template for data entry.

The DC fields also had to be mapped correctly to the OPAC search function. Library catalogues allow users to search for material by title, author, subject, and keyword. Users may also limit their searches by specifying other fields, such as publication date. Which DC fields should be mapped to which search fields? Should article title and journal title both be title searches? Should publication date be the date the electronic copy was created or the date the original article was published or both? Many decisions, once made, are not easily undone after hundreds of records are indexed. These indexing decisions all had to be made before the DC cataloguing for the *AHJ* articles could begin.

Finally, in March 2005, data entry using qualified DC began in earnest. All *AHJ* articles from 1974 through 1992 had bibliographic records using qualified DC by June 2005.

The final piece of the puzzle is presentation. A public interface was needed to guide users through the digital collection. First, there had to be a home web page for the DAC as a whole, with appropriate search engines and explanations embedded. Then there had to be web pages for each of the sub-collections and intermediate pages to display search results. A final set of pages was needed to display the individual catalogue records with links to the full-text PDF files, the files containing article page images, and the JPEG files for the issue covers. Microsoft FrontPage and Macromedia Dreamweaver software were used to create the web pages for the DAC.

ADDING OTHER COLLECTIONS

The core of the university's Digital Accounting Collection, the first 19 volumes of the *Accounting Historians Journal*, was now in place. It was time to add other material. The library and the AICPA had long been looking at a cooperative effort to digitize some of the Institute's materials. After discussing several ideas at the semi-annual AICPA-UML meetings, committee members agreed that the AICPA collection of non-current exposure drafts should be added. This collection consists of some 350 documents, the first drafts of AICPA accounting and auditing standards which were printed and circulated for comment in

advance of issuance as a final, authoritative standard.

The exposure drafts are not widely held since they were never meant to be part of the permanent record. This makes them a particularly useful item to add to an online research collection. The drafts range in time from 1968 through 2004. The collection is limited to non-current exposure drafts; that is, those no longer under discussion because a final standard was issued or the exposure draft was withdrawn from consideration. The exposure drafts represent an important research resource for historians and practitioners exploring the development of standard setting. Currently, only the page-image PDFs of the exposure drafts are being displayed in the DAC.

The third group of materials identified for digitization is the Accounting Pamphlet Collection. The AICPA holdings at the University of Mississippi include thousands of accounting pamphlets, hundreds of which date from the early 1900s (hence, are in the public domain). These pamphlets were published as practical guides for bookkeepers and accountants. They represent valuable insights into the operations of late 19th and early 20th century businesses, industries, and service organizations. Many are unique items held only in this collection. The digital pamphlet collection is currently not large, comprising a little over one hundred pamphlets that have been scanned, digitized, and catalogued using DAC. These range from ten to fifty pages in length and cover the period 1905 to 1924.

Occasionally, in working with a collection, fun items pop up and simply demand to be shared. In the process of creating display cases and brochures for the accounting collection, many images from frontispieces, magazine mastheads, old group photographs, and other illustrations have been located and electronically scanned. As an offshoot of this work, the Accounting Photographs and Images Collection emerged, a small selection of accounting and bookkeeping illustrations from across the ages. Suggestions for other illustrations are welcome.

The DAC now has: (1) the *Accounting Historians Journal* Collection, volumes 1-19 (a completed collection of 450 articles), (2) the AICPA Exposure Draft Collection (eventually about 350 documents, though not all are yet available in the digital collection), (3) the Accounting Pamphlet Collection, and (4) the Accounting Photograph and Images Collection. These collections are all managed and searched by the III software that is used to run major library catalogues around the world. The elapsed project time, from the point students were hired to begin scanning and digitizing the material until the website was publicly

functional, was a little over two years. If the library staff had learned what they needed to know before beginning the project, and if the necessary software had already existed and been debugged, it would probably have taken less than half that time.

THE FUTURE

It has been an eventful five years since James Davis and Dale Flesher walked over to the UML with the AICPA's request for a proposal in hand. Little did anyone realize that accounting history would propel the UML into the realm of digital collections. The UML, already the home of the National Library of the Accounting Profession, has also become the site of a growing Digital Accounting Collection, built on the foundation of *AHJ*.

The ongoing work already outlined will keep the librarians busy for quite a while, but it is always fun to dream of other projects. Not all of these need be purely historical. *The Accounting Historians Notebook* is an obvious contemporary choice. The library's association with the AICPA will also provide opportunities to digitize important documents illustrating the Institute's work and history. Publications from the early years of other accounting organizations, such as the National Association of Cost Accountants, should be on a list of digital projects as well. Furthermore, the library has received the early work papers of Haskins and Sells. Digital images of many of these papers have already been made. Adding them to the DAC would provide access to manuscript materials in the field. Finally, suggestions by users are always the most valuable input as to new items to digitize. Their suggestions will ensure that historical documents important for scholarly research are made accurately and easily accessible online.

REFERENCES

- An Act Making Appropriations for the Departments of Labor, Health and Human Services, and Education, and Related Agencies for the Fiscal Year ending September 30, 2002, and for Other Purposes.* (January 10, 2002), Public Law 107-116, *U.S. Statutes at Large* 115:2177.
- Anonymous (1998), "An Institution at the Institute," *Journal of Accountancy*, Vol. 185, No. 3: 98.
- Anonymous (2001), "AICPA Collection Transferred to University of Mississippi, Member Services Continue," *The CPA Letter*, Vol. 81, No. 7: 2.
- John Davis Williams Library (2005), *University of Mississippi Users' Guide for Dublin Core Metadata Cataloging* (University, MS: J.D. Williams Library).
- Martin, M.M. (2004), "Adapting Reference for a Unique Group of Distance Learners: Serving the American Institute of Certified Public Accountants (AICPA)," *Journal of Library and Information Services in Distance Learning*, Vol. 1, No. 4: 59-66.

- Neloms, K.H. (1987), "History of the AICPA Library," *Journal of Accountancy*, Vol. 163, No. 5: 388-392.
- Rothberg, J.L. (2000), letter sent to J.W. Davis, August 18, 2000.
- Samson, W.D (2002), letter sent to R. Kurtz, November 25, 2002.
- University of Mississippi (2000), *Plan for a National Library of the Accounting Profession: A Proposal to the American Institute of Certified Public Accountants* (University, MS: University of Mississippi).

APPENDIX A

Acronyms Used

AAH	Academy of Accounting Historians
AHJ	<i>Accounting Historians Journal</i>
AICPA	American Institute of Certified Public Accountants
DAC	Digital Accounting Collection
DC	Dublin Core
III	Innovative Interfaces, Inc.
IMLS	Institute of Museums and Library Services
MB	Metadata Builder
OCR	optical character recognition
OPAC	online public access catalogue
UML	University of Mississippi Library