Southern Illinois University Carbondale
OpenSIUC

Articles | Morris Library

3-2007

On Your Mark, Get Set, Go! Overview of a Digital Project from Start to Finish

Pam Hackbart-Dean
Southern Illinois University Carbondale, phdean@lib.siu.edu

Doug Goans
Georgia State University

Lauren Kata
Applied Engineering Solutions, Inc.

Follow this and additional works at: http://opensiuc.lib.siu.edu/morris_articles
Published in *Computers in Libraries*, Vol. 27, No. 3 (March 2007).

Recommended Citation
Hackbart-Dean, Pam, Goans, Doug and Kata, Lauren. "On Your Mark, Get Set, Go! Overview of a Digital Project from Start to Finish." (Mar 2007).

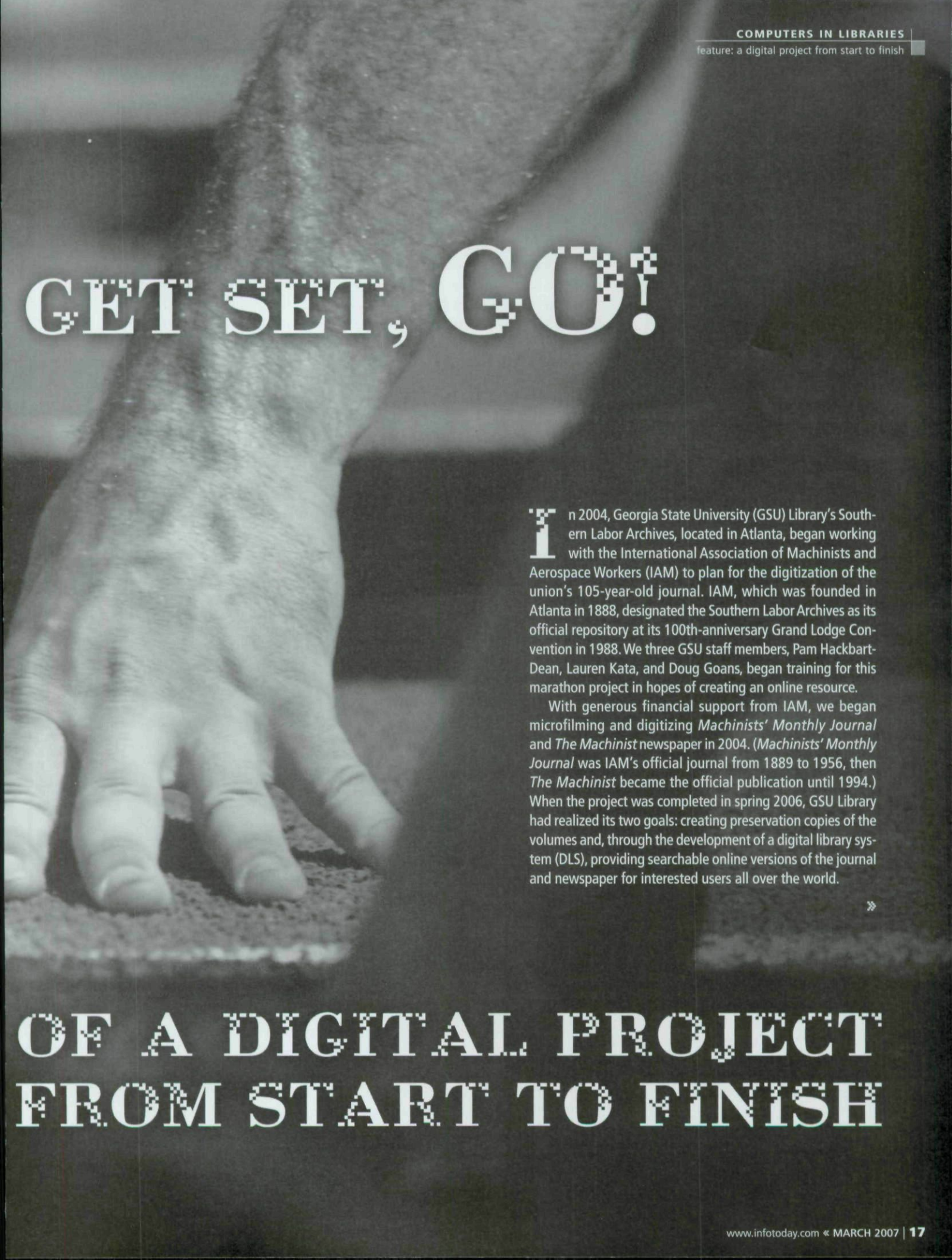
This Article is brought to you for free and open access by the Morris Library at OpenSIUC. It has been accepted for inclusion in Articles by an authorized administrator of OpenSIUC. For more information, please contact opensiuc@lib.siu.edu.

On your mark,

BY DOUG GOANS, PAM HACKBART-DEAN, AND LAUREN KATA

When this marathon project was completed, GSU Library had realized its two goals: creating preservation copies of the volumes and, by developing a digital library system, providing searchable online versions of these publications for interested users all over the world.

OVERVIEW



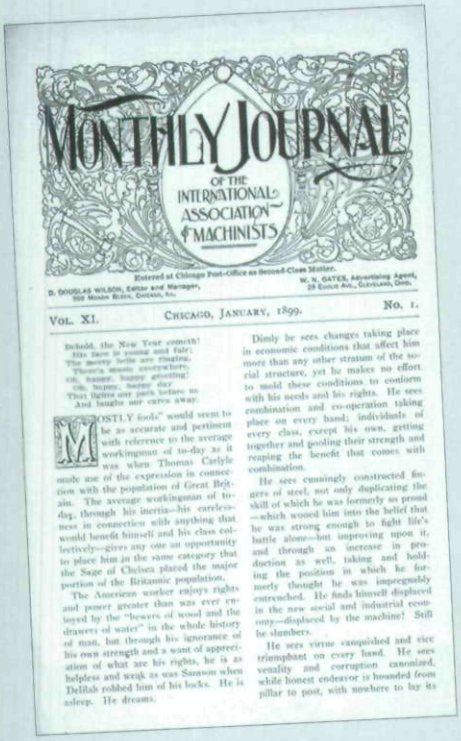
GET SET, GO!

OF A DIGITAL PROJECT FROM START TO FINISH

In 2004, Georgia State University (GSU) Library's Southern Labor Archives, located in Atlanta, began working with the International Association of Machinists and Aerospace Workers (IAM) to plan for the digitization of the union's 105-year-old journal. IAM, which was founded in Atlanta in 1888, designated the Southern Labor Archives as its official repository at its 100th-anniversary Grand Lodge Convention in 1988. We three GSU staff members, Pam Hackbart-Dean, Lauren Kata, and Doug Goans, began training for this marathon project in hopes of creating an online resource.

With generous financial support from IAM, we began microfilming and digitizing *Machinists' Monthly Journal* and *The Machinist* newspaper in 2004. (*Machinists' Monthly Journal* was IAM's official journal from 1889 to 1956, then *The Machinist* became the official publication until 1994.) When the project was completed in spring 2006, GSU Library had realized its two goals: creating preservation copies of the volumes and, through the development of a digital library system (DLS), providing searchable online versions of the journal and newspaper for interested users all over the world.

»



MONTHLY JOURNAL
OF THE
INTERNATIONAL ASSOCIATION
OF MACHINISTS

Entered at Chicago Post-Office as Second-Class Matter.

D. DOUGLAS WILSON, Editor and Manager, 708 Monon Block, Chicago, Ills.

W. N. GATES, Advertising Agent, 29 Euclid Ave., Cleveland, Ohio.

VOL. XI. CHICAGO, JANUARY, 1899. NO. 1.

Behold, the New Year comes!
He here is coming and fair;
The merry bells are ringing,
There's music everywhere;
Oh, happy, happy greeting!
Oh, happy, happy day!
That lights our path before us
And laughs our cares away.

MOSTLY fools" would seem to be as accurate and pertinent with reference to the average workingman of to-day as it was when Thomas Carlyle made use of the expression in connection with the population of Great Britain. The average workingman of today, through his inertia—his carelessness in connection with anything that would benefit himself and his class collectively—gives any one an opportunity to place him in the same category that the Sage of Chelsea placed the major portion of the Britannic population.

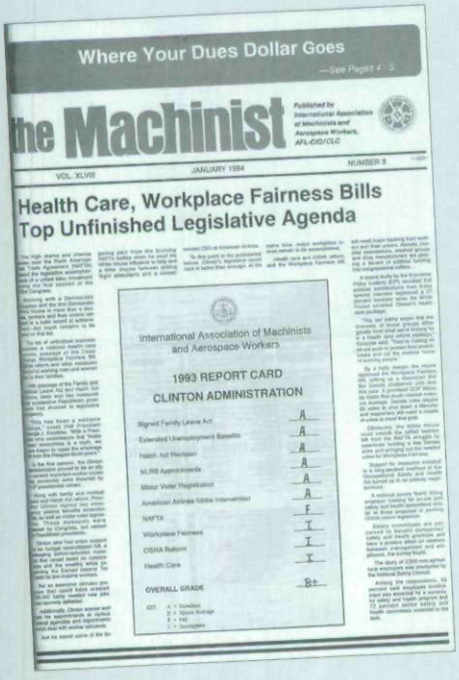
The American worker enjoys rights and powers greater than was ever enjoyed by the "hewers of wood and the drawers of water" in the whole history of man, but through his ignorance of his own strength and a want of appreciation of what are his rights, he is as helpless and weak as was Samson when Delilah robbed him of his locks. He is asleep. He dreams.

Dimly he sees changes taking place in economic conditions that affect him more than any other stratum of the social structure, yet he makes no effort to mold these conditions to conform with his needs and his rights. He sees combination and co-operation taking place on every hand; individuals of every class, except his own, getting together and gauging their strength and reaping the benefit that comes with combination.

He sees cunningly constructed fingers of steel, not only duplicating the skill of which he was formerly so proud —which wooed him into the belief that he was strong enough to fight life's battle alone—but improving upon it, and through an increase in production as well, taking and holding the position in which he formerly thought he was impregnably entrenched. He finds himself displaced in the new social and industrial economy—displaced by the machine! Still he slumbers.

He sees virtue vanquished and vice triumphant on every hand. He sees venality and corruption canonized, while honest endeavor is hounded from pillar to post, with nowhere to lay its

Machinists' Monthly Journal *was first published in 1889.*



Where Your Dues Dollar Goes
—See Pages 4 - 5

the Machinist

Published by International Association of Machinists and Aerospace Workers, AFL-CIO/CLC

VOL. XLVII JANUARY 1994 NUMBER 2

Health Care, Workplace Fairness Bills Top Unfinished Legislative Agenda

International Association of Machinists and Aerospace Workers

1993 REPORT CARD
CLINTON ADMINISTRATION

Signed Family Leave Act	A
Extended Unemployment Benefits	A
Hatch Act Reform	A
WARN Appointments	A
Motor Voter Registration	A
American Airlines Strike Intervention	F
NAFTA	I
Workplace Fairness	I
OSHA Reform	I
Health Care	I
OVERALL GRADE	B+

The Machinist was published in a newspaper format beginning in 1946.

IAM's Project Plan

Our major objective for this digital conversion project was to improve access to the publications for IAM members as well as for scholars, historians, students, and the general public. People's use of this collection would no longer be limited by geography or time.

Machinists' Monthly Journal and *The Machinist*, which contain images of and stories about American and Canadian workers, documented activities of IAM and its local lodges for more than a century. Ultimately, these journals provide a unique snapshot of the union and the changing roles of its members in the workplace and society.

The project's plan was to convert the publications into microfilm and then scan these new images into digital image files. These files would be further processed with additional optical character recognition (OCR) software to create full-text versions of the content. The resulting digitized material would then be searchable via the Web.

The following personnel were involved in the project:

- Pam Hackbart-Dean, head of special collections and archives, was the project director. She managed the overall project and served as the point person for vendors.
- Lauren Kata, archivist for the Southern Labor Archives, consulted on materials management, workflow coordination, quality-control checking, and usability testing.
- Doug Goans, Web development librarian, managed interface development and the construction of the in-house DLS.
- Jackey Hong, a programmer in the systems office, implemented and coordinated metadata and built the infrastructure for the in-house DLS.
- Amanda Brown, a full-time library technical assistant in the special collections department, conducted quality-control checking and assisted with usability testing throughout the project. Fifty percent of her time was dedicated to this project.
- Casey Cater, a part-time graduate assistant in the special collections department, contributed to initial quality-control checking.

Thus, our in-house team consisted of five full-time employees and one part-time employee. Our need to balance time and responsibilities between this project and our primary duties contributed to the project's 2-year length.

Training for the Race

The primary benefits of outsourcing digital conversion work to a vendor are financial and technical. Keeping up with current software, equipment, and trained personnel is the vendor's responsibility. This meant that we wouldn't have to provide physical space for scanners or for those folks who were doing the scanning.

Identifying and choosing a vendor is not quick or easy, especially for large and complex projects. The vendor selection process involves the following:

- Developing an initial concept of the project and its goals
- Identifying potential vendors
- Sending out an RFI (request for information) to explain the project's goals clearly and to discover which vendors are interested and have ideas about how to handle it
- Establishing a project methodology and quality requirements
- Developing a short list of vendors

- Writing an RFP (request for proposal) and sending it to those that are on the short list along with samples to be scanned
- Communicating with the vendors while they work on their responses, including meetings when possible
- Evaluating and comparing the vendors' proposals and selecting the best one
- Writing and signing a contract[1]

We knew we wanted a system that would deliver a usable and user-friendly interface, provide record management and maintenance, and offer compatibility with our Web services infrastructure. As a result, we developed the following list of questions:

1. Do you provide PDF and text files?
2. What is your storage medium?
3. Could you estimate the storage size for this project?
4. What would be the scale for the individual PDF?
5. Do you offer color, grayscale, and/or bitonal?
6. Will you provide single page or bound PDF?
7. Do you provide TIFF, JPEG, or GIF files?
8. What is your percentage of accuracy for OCR?
9. Will you provide metadata and let us see a sample?

Not all vendors answered the above questions the same. Nonetheless, all were able to provide PDF and text-only files as well as offer DVD as a storage medium. (However, estimates for storage space varied among vendors.) Each vendor also reported it would digitize at a minimum resolution of 300 dpi in grayscale. Only one, OCLC, offered to provide both a single page and bound PDF file of a single journal.

About the time we had finished assessing our digital library system (discussed below), we selected OCLC as our digitization vendor. OCLC provided us with options for the types of deliverables we would need. We wanted PDF files for the online collection both for searching and display. We selected the text files so that we could have pure text versions of the archive for searching and also to accompany the pure TIFF images in the event that we would ever need to rebuild a PDF in the future. In short, we now have pure images of the documents as they looked historically as well as separate OCR text files of the documents. OCLC provided an Access database of metadata that also gave us some relief in the building of the system: We would be able to harvest the information to build a back-end database in MySQL to support search, browse, and display functions.

Warm-Up Exercises for Research and Development

Both during and after the search for a digital conversion vendor, we planned for how the resulting digital files would be stored, disseminated, and searched. This wasn't a traditional Web page publishing project. We would need a digital library system with navigation, search, and display mechanisms specific to this collection.

As with previous projects undertaken at Georgia State University Library, we assessed the pros and cons for different digital library systems. The options that emerged were similar in nature to other Web projects. We could build something from scratch, find a "free" open source solution, look for a hosted solution from a vendor, or purchase a product from the vendor to locally install.

Up to this point, the library had not implemented any formal DLS applications but instead had published collections online, including archival photographs and finding aids. We had not invested in a dedicated DLS built around metadata standards that would provide compliance for emerging archival, harvesting, and searching technologies. Implementing something of this scale was a large task, and the realities of tight budgets and small staff sizes were looming.

"Ultimately, these journals provide a unique snapshot of the union and the changing roles of its members in the workplace and society."

We were aware of DLS solutions chosen by other institutions. We considered commercial products, specifically Olive and CONTENTdm. However, the library deemed their costs prohibitive for this initial project. We subsequently tested the free, open source product Greenstone. We compared Greenstone's application code and requirements for customizing the system with our own Web content management system that was being built in-house using ASP and a MySQL database. While Greenstone appealed to us initially, we determined that the time and learning curves required to support local customization of the system were unsuitable to meet the project's timeline. In addition, it wouldn't easily coexist with the development platform (specifically ASP and PHP) we utilized regularly for other Web projects.

We ultimately chose to pursue an in-house DLS as it would enable us to maintain influence over the look and

»

feel as well as support anticipated value-added features such as tutorials, instruction, and links to related resources. We decided that for this initial digital library project we could approximate the same basic search and browse features of Greenstone or other digital library systems using our existing Web development infrastructure and programming skills.

Off and Running with Implementation and QC

The evaluations described above occurred roughly between June 2004 and February 2005. Once we'd selected OCLC as our vendor and decided to create the DLS in-house, we designed a workflow plan. The project director assembled the materials and shipped them to OCLC between September and October 2004.

OCLC microfilmed and digitized the materials between September 2004 and March 2005. It delivered digital copies on DVDs and external hard drives approximately every 20 days. The programmer began DLS design, and back-end development commenced as we received digital files. (This all culminated in a final launch in May 2006.)

As files came to us from OCLC, the Web development librarian transferred the PDF files to our in-house Web server and imported the metadata from the Access databases into MySQL. We used Active Server Pages to build the interface that communicated with the MySQL database. This interface would allow users to browse the collection by issue or year, with links to the PDF documents that displayed the full-image view of the publication. We soon began to work with the campus IT department to enable full-text searching of the collection using the campus Google search appliance.

Our teams met regularly to review the developing site's progress. As we did, we began noticing inconsistencies. For example, in some cases the OCR text was not matching up with the PDF files because the Access database tables containing metadata would sometimes have inconsistencies in the field names from batch to batch. To correct this, we needed to locally adjust the metadata deliverables that OCLC had sent us. There were also errors in the OCR scanning that sometimes made searching dubious. These inconsistencies made it clear that we needed to build a quality-control (QC) process into our workflow.

"We found the quality-control process to be a very necessary part of the project."

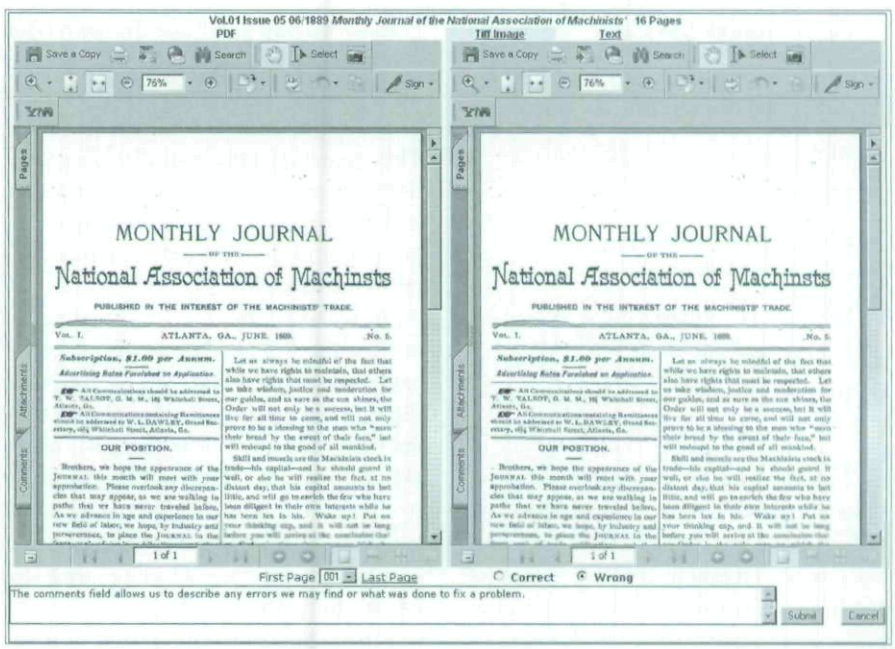
Our programmer created a quality-checking system so that our library technical assistant and graduate assistant could review the digital files. This review included checking PDF files against text and TIFF files as well as looking for any glitches or other inconsistencies. We found the quality-control process to be a very necessary part of the project.

Finish Line in Sight

Periodic demonstrations of the site aided in our ongoing assessment and evaluation of project needs. Our first public demo took place in February 2005 at a luncheon for the Georgia Council of Machinists. At this event, we were able to provide a visual progress report to our funding organization and also to stimulate the interest of its local members, whom we saw (and continue to see) as the project's primary beneficiaries.

In an evaluation meeting following the first public demo, we addressed questions of branding, access, copyright, and usability. Although we knew the goal was to make the collections accessible worldwide, we also knew that

»

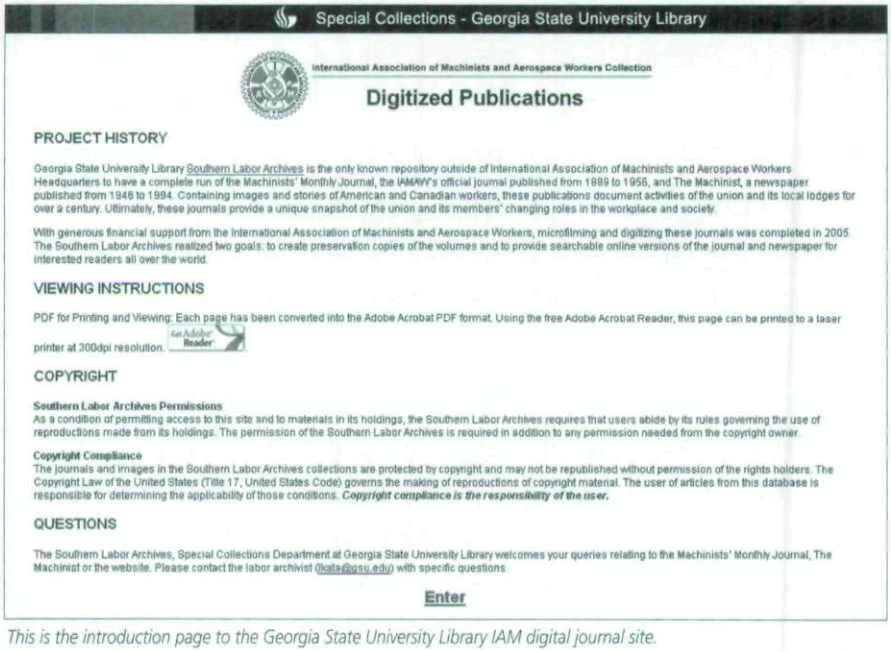


Quality checking enabled us to review the digital files via a Web browser and identify and correct problems along the way, such as missing or inconsistent metadata that rendered PDF files either invisible or inaccurate. Quality checking involved viewing sample pages from each PDF file (both the journal and newspaper), comparing with the TIFF and OCR text version of that page, and marking whether there were any problems or if the file was accurate.

copyright was still held by IAM and that users would have to be made aware of the copyright restrictions.

We decided that since there were more than 70,000 pages, it would be unrealistic to add a copyright branding to each one. In evaluating other institutions' notification techniques, we found that some sites prompt users to agree to copyright terms when accessing copyrighted content. We rejected that option. Ultimately, we decided to include an introductory paragraph about copyright on the main page and to add an additional copyright notice to the beginning of each PDF file.

Our steady workflow, the addressing of copyright and branding issues, and the final implementation of the interface continued through November 2005. By December 2005, we were ready to assess the usability of our product by collecting feedback in three general areas: cross-browser support, using PDF Reader software, and using the Google search engine that we applied to our interface. The usability assessment began first among ourselves and fellow co-workers and was later conducted through one-on-one interviews with selected colleagues and union member focus groups.



Special Collections - Georgia State University Library

International Association of Machinists and Aerospace Workers Collection

Digitized Publications

PROJECT HISTORY

Georgia State University Library Southern Labor Archives is the only known repository outside of International Association of Machinists and Aerospace Workers Headquarters to have a complete run of the Machinists' Monthly Journal, the IAMAW's official journal published from 1889 to 1956, and The Machinist, a newspaper published from 1946 to 1994. Containing images and stories of American and Canadian workers, these publications document activities of the union and its local lodges for over a century. Ultimately, these journals provide a unique snapshot of the union and its members' changing roles in the workplace and society.

With generous financial support from the International Association of Machinists and Aerospace Workers, microfilming and digitizing these journals was completed in 2005. The Southern Labor Archives realized two goals: to create preservation copies of the volumes and to provide searchable online versions of the journal and newspaper for interested readers all over the world.

VIEWING INSTRUCTIONS

PDF for Printing and Viewing: Each page has been converted into the Adobe Acrobat PDF format. Using the free Adobe Acrobat Reader, this page can be printed to a laser printer at 300dpi resolution. Get Adobe Reader

COPYRIGHT

Southern Labor Archives Permissions
As a condition of permitting access to this site and to materials in its holdings, the Southern Labor Archives requires that users abide by its rules governing the use of reproductions made from its holdings. The permission of the Southern Labor Archives is required in addition to any permission needed from the copyright owner.

Copyright Compliance
The journals and images in the Southern Labor Archives collections are protected by copyright and may not be republished without permission of the rights holders. The Copyright Law of the United States (Title 17, United States Code) governs the making of reproductions of copyright material. The user of articles from this database is responsible for determining the applicability of those conditions. ***Copyright compliance is the responsibility of the user.***

QUESTIONS

The Southern Labor Archives, Special Collections Department at Georgia State University Library welcomes your queries relating to the Machinists' Monthly Journal, The Machinist or the website. Please contact the labor archivist (lkata@gsu.edu) with specific questions.

Enter

This is the introduction page to the Georgia State University Library IAM digital journal site.

Note: Scroll down to view the issue.

Georgia State University Library
Special Collections Department
Southern Labor Archives

Copyright Notice

This digital collection has been made available courtesy of Georgia State University, with support from the International Association of Machinists and Aerospace Workers (IAMAW).

Copyright Compliance
The journals and images in the Southern Labor Archives collections are protected by copyright and may not be republished without permission of the rights holders. The Copyright Law of the United States (Title 17, United States Code) governs the making of reproductions of copyright material. The user of articles from this database is responsible for determining the applicability of those conditions. ***Copyright compliance is the responsibility of the user.***

Here's the copyright notice that's branded on every individual digitized issue.

We found that many of the usability issues or concerns were related to the individual user's familiarity with his or her browser, the Google interface, and PDF Reader software. To address these issues, we have begun to develop basic and advanced online instruction tutorials and plan to continue with usability testing. Our programmer handled minor glitches that people found during the usability assessment, such as the lack of browser standardization and pop-up blocking. In all, we microfilmed and digitized approximately 71,000 pages of *Machinists' Monthly Journal* and *The Machinist*.

In May 2006, the project director flew to Hollywood, Md., to unveil the completed online resource to IAM. "This project uses the latest technology to take us back to our earliest days. We want our members to use this site and to learn about our past and our struggles so we can stay strong in the future," said Bob Thayer, IAM general vice president, who was at the debut.[2] Since May 2006, when the Web site went live, *Machinists' Monthly Journal* and *The Machinist* have been available for the entire world to use.

Crossing the Finish Line

As we continue to collect user feedback on the Web site's functionality, we have also received enthusiastic responses about the value of the online collection. "This has been a tremendous project to capture our (IAM) history and make it available to our members, labor scholars, and anyone interested in labor history," said one IAM union member. Another user, a genealogist who located information about her grandfather, described the online collection as "incredible." She said, "These publications give not only the IAM's history but are a source for genealogists, scholars, and students to get labor's perspective on some of the greatest events in North American history." Since we launched this site, it has had 355,526 hits, 273,546 document views, 99,949 visits, and 65,655 unique visits. You can visit the site at www.library.gsu.edu/dlib/iam.

We hope to be able to re-evaluate our solutions and consider whether we will need to either revisit an open source solution or talk at a higher level about investing in a commercial product to manage our digital assets, which continue to grow in special collections. Another option might be to collaborate with other institutions to implement a

consortium product for managing our collective digital assets.

The costs for this project were for microfilm and digital conversion services, which we outsourced to a vendor, and for DLS development and Web design, which was handled in-house. Ongoing costs (absorbed by the library) will include Web site development and maintenance.

We think it's important to have a conscious, planned commitment to preserve the information from our project at an institutional level.

As a final point, we think it's important to have a conscious, planned commitment to preserve the information from our project at an institutional level. Commitments such as permanent assignment of staff, designated responsibilities, and recognition that electronic preservation will require specific technical experience in areas of metadata, archiving, and digitization are ways that Georgia State can support digital library collections. Additionally, the permanent existence of any digital library collections and initiatives will require guaranteed operational funding, maintenance and migration of data, and related technical care.

Ultimately, we realized our goals of creating preservation copies of the journal and newspaper volumes and providing searchable online versions of both for interested users all over the world. We are very satisfied to have gone the distance in this marathon.

Doug Goans is the Web development librarian at Georgia State University Library in Atlanta. He holds an M.L.I.S. from the University of Missouri–Columbia. His email address is dgoans@gsu.edu. At the time of this project, Pam Hackbart-Dean was head of special collections and archives at Georgia State University. She is now the director of the Special Collections Research Center at Southern Illinois University–Carbondale. She holds an M.A. from the University of Connecticut in Storrs, Conn. Her email address is phdean@lib.siu.edu. At the time of this project, Lauren Kata was the archivist for Georgia State University's Southern Labor Archives. She is now an archivist at Applied Engineering Solutions, Inc. in Tampa, Fla. She holds an M.A. in history and a graduate certificate in archival administration from Wayne State University in Detroit. Her email address is lauren.kata@gmail.com.

Endnotes

1. Janet Gertz, "Vendor Relations," *Handbook for Digital Projects: A Management Tool for Preservation and Access*. (Andover, MA: Northeast Document Conservation Center, 2000): 144.
2. "IAM History Makes Online Debut," iMail, 23 May 2006, www.iamaw.org/publications/imail/imail_05_23_2006.htm (06 June 2006).

A WEALTH
of marketing information and tips for you and your library

Marketing Library Services
The "How-To" Marketing Tool
Written Exclusively for Librarians

The Marketing Library Services newsletter is the premier source of ideas and news for savvy library marketers. Whether you work in a corporate, special, academic, school, or public library, MLS will help you call attention to your library, your services, and your value to the community you serve.

MLS will provide you with valuable how-to information on developing marketing strategies, creating marketing materials, organizing traffic-generating programs, acquiring library funding, and so much more.

Reserve your FREE copy of MLS today

For a FREE TRIAL ISSUE of **MLS** simply call (800) 300-9868 or send an e-mail to custserv@infotoday.com.

If you like it, pay just $73.95 for five more issues (6 issues total) when the bill arrives. That's a savings of $10 off the basic U.S. subscription price. If you are not pleased with your FREE copy, return the bill marked "Cancel" and have no further obligation.

www.infotoday.com

Information Today, Inc.
143 Old Marlton Pike
Medford, NJ 08055

Copyright of Computers in Libraries is the property of Information Today Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.