FINAL REPORT
TASK FORCE ON ELECTRONIC PUBLICATIONS
SOCIETY OF AMERICAN ARCHIVISTS

Robert P. Spindler, chair
Nancy Bartlett (American Archivist Editorial Board)
Teresa Brinati (SAA Publications Director)
Robin Chandler
Solveig DeSutter (SAA Education Office)
Brian Doyle (SAA Webmaster)
Philip B. Eppard (American Archivist Editor)
Susan Fox (SAA Executive Director)
Ed Galloway
David Haury (Council Liaison)
Cal Lee
Dennis Meissner (SAA Publications Board)
Patti O'Hara (SAA Education Office)
Johanne Pelletier
Scott Schwartz

TABLE OF CONTENTS

Charge:

On December 26th, 2001 SAA Council charged the SAA Task Force on Electronic Publishing (TFEP), "To devise a plan for phased electronic publication of periodicals and monographs produced by the Society of American Archivists." The charge identified several questions to be addressed by TFEP:

- How can SAA use electronic publishing to further its strategic goals?
- Given SAA's mission, what are the needs of its members with respect to electronic products?
- How can electronic publishing be used to decrease the cost of creating and/or delivering content?
- How does SAA view electronic publishing? (e.g. as a service to its members? Publishing profit center? An information tool? Marketing/communication vehicle?)
- How can we determine what effect electronic publishing will have on the bottom line?
- When do we need to protect SAA content and rights?
- What copyright issues might be involved in converting past issues of SAA publications, and how does SAA feel about them?

Process:

SAA provided support for the task force to hold a kickoff meeting on a snowy weekend in Chicago on March 1-3, and the formal TFEP members were asked to complete a pre-meeting assignment to survey several e-publishing organizations and their practices. "Archival values and electronic publishing" was the assigned dinner discussion topic for Friday night, which led to an archival values exercise Saturday morning in which the group brainstormed responses to the questions in the TFEP charge. The group then heard the pre-meeting assignment reports, and Teresa Brinati concluded the morning work with a review of past SAA work on publishing issues. The dinner discussion, pre-meeting reports and the values exercise served as an environment scan to provide context for our deliberations.

On Saturday afternoon we planned to develop charges for two subcommittees, technical and business model, but the group wanted to discuss specific outcomes and products. We spent much of Saturday afternoon discussing our perceptions of SAA membership needs, existing and potential new products, and the concepts of aggregation and disaggregation. We ultimately focused on discussing the workflow for managing discrete items of content associated with different SAA titles, and this discussion continued into Sunday morning. As a result of the content flow discussion we decided to establish a content subcommittee and a technical subcommittee because it appeared that we needed to have a vision for what we wanted to do in place before we could really address technical and business model issues in depth. Individuals volunteered to write vision statements for particular types of content, and others volunteered to write documents regarding four technical issues: submission, retroconversion, preservation and infrastructure. The content experts were charged to bring their reports to the technical experts first and the technical experts then commented on the feasibility of the content recommendations, incorporating those ideas into their technical reports. These documents are appended to this report.

A TFEP progress report was submitted to Council in May 2002, and comments from Council were received and incorporated into an "Exposure Draft", completed August 9, 2002. In the Exposure Draft several of the Appendices were revised to include more specific business model information, the Preservation Report was expanded and the Infrastructure Report was added to this edition. The Exposure Draft was posted on the SAA website ca. August 15, 2002 and hardcopy was distributed at the SAA 2002 Annual Meeting in Birmingham, Alabama. The TFEP chair hosted office hours at the Annual Meeting. Few comments were directly received from membership although substantive comments were received from Richard Cox and from the Council of State Historical Records Coordinators, who cited the May edition of this report in their document *Connecting the Archival Community: Report of the Archival Education and Information Web Needs Assessment Project* (July, 2002 www.coshrc.org/reports) An October draft of the TFEP report was shared with the SAA Membership Committee and the Publications Board for comment. No comments were received from those bodies, although TFEP member Scott Schwartz worked with the Membership Committee to create Appendix XII, a list of draft questions for a possible future membership survey. The Appendixes of this report (other than Appendix XII and the retroconversion costs added to Appendix VIII)

have not been modified to account for TFEP discussions in November, 2002, so there may be some differences between the recommendations stated in the body of the report and those previously stated in the Appendixes.

Content Vision Summary:
Probably the most powerful concept that emerged from these discussions was the idea of aggregating and disaggregating content, or repurposing chunks of content to meet different presentation, delivery and information needs of SAA and its membership. TFEP envisions accepting content submissions through a single website and completing licensing agreements with authors that enable SAA to reuse or repurpose content to suit its current and future needs. For example, an article submitted for the *American Archivist* could be repackaged and distributed as part of a CD ROM publication of information about managing email that contains an Annual Meeting presentation, news notes from *Archival Outlook* about related projects at institutions, and perhaps an item from a proposed new series of technical leaflets. We also discussed the possibility of creating shopping cart functionality on the SAA website that would enable users to self-select their own combinations of content pieces, their preferred delivery vehicle (i.e. email attachments, CD or s-mail hardcopy) and method of payment. Electronic delivery of content might be provided at a discount compared to hardcopy delivery. We believe that international accessibility to self-selected combinations of content or SAA-repackaged units of content may be the greatest sales growth opportunity for SAA, and a substantial enhancement of member benefits in terms of potential discounts for electronic delivery, increased variety, timely access and international recognition of their writing.

Technical Vision Summary:
Several recommendations have been proposed regarding retroconversion of the *American Archivist* (See Appendix VIII):

- Acquire complete run of issues from either the SAA Archives or volunteers
- Disbind for production efficiency and scanning accuracy
- Scan issues at 600 dpi to make TIFF 6.0 (1-bit) archive image files
- Convert to GIF on the fly for web presentation
- Enable printer-friendly PDF version
- OCR for full text searching but do not correct errors, display only GIF images
- Create several forms of metadata that support internal needs and OAI
- Develop RFP and outsource the work
- Outsource hosting, serving and maintenance
- Seek future SGML/XML solution for preservation

TFEP also proposes the creation of a single website for submission of all SAA content (see Appendix IX), and eventual creation of online editing and review processes.[1] We believe that descriptive and administrative metadata can be supplied by authors or captured during the submission process that will facilitate subsequent use, reuse and preservation of this content. Appropriately maintained content file format (and soon XML markup) submission standards will facilitate reformatting of content to production, distribution and preservation formats. The rigor of the standards needs to be balanced with the technical expertise of our authors to facilitate rather than retard submissions. Member benefits would include more timely editorial, publication and distribution functions and privileged advance access to certain content for members only (delayed release for non-members).

TFEP also believes that while certain production and distribution costs may be reduced through technology, the costs for maintaining and preserving these products represents the largest unknown element of any e-publication business plan. To date we have not identified any reliable data regarding long-term

---

[1] The interface design presented in Appendix IX was intended to accommodate all SAA publications, but Brian Doyle notes in his Infrastructure Report (Appendix XI) that it may be desirable to have a single submission *website* with several submission *interfaces* customized for the nature of the content, since some content will be submitted as attachments and other shorter items might be submitted by pasting text into a text box.

maintenance, migration and repair costs for electronic publications. The short-term data that is available may not be reliable due to the very small sample and our shallow national and international experience in this area.

Given the uncertain future of specific technologies it is likely that SAA should hedge its bets by retaining and maintaining preservation copies of digital objects in more than one format[2] These measures may not be appropriate in all cases, since some of our electronic content may not warrant long-term preservation. Preservation solutions may need to be sought for both the discrete content "chunks"and for their presentation in the context of a formal publication. Length of retention may differ depending on whether the content is permanently retained in hardcopy and whether the electronic version is likely to be needed for reaggregations, repurposing, or administrative functions. As a result, version control will be a significant challenge in this environment. Preservation and administrative metadata should be created and retained with the preservation versions of content. Preservation and accessibility issues should be contractually addressed with any external service provider.

Next Steps, Prioritized:

TFEP discussions of these priorities revealed differences of opinion in how quickly various initiatives could be implemented. There was substantial interest in quickly moving forward on changes to the *American Archivist*, *Archival Outlook* and SAA hosting and retention of Annual Meeting presentations and speaker biographies. This desire to move quickly was tempered by other TFEP members concerns for the availability of resources to accomplish these ends. As a result TFEP recommends establishment of a Working Group on Electronic Publications hosted by the SAA Publications Board to work with SAA Council and SAA office staff to identify available resources and to refine and implement the following priorities.

Some existing and related technology initiatives of the SAA Office were brought to the attention of the committee and have been added to the priorities for 2003-2004 in order to give a fuller picture of the current obligations of the SAA office staff. Several TFEP members called for market research to identify our members' interests and perceptions of e-publishing. As a result the priorities tend to delay some e-publishing initiatives until market research can be completed, meanwhile a substantial slate of infrastructure and legal issues can be addressed. Other simple versions of e-publishing might be implemented in the first two year cycle while more sophisticated implementations are delayed (e.g. posting PDF versions of *Archival Outlook* after a three-month delay, which enables the intended temporary embargo w/o member authentication) TFEP member Scott Schwartz has worked with the SAA Membership Committee to draft several related questions that might be used in a future member survey. These questions have been added to this report in Appendix XII.

Following are suggested priorities divided into three potential time frames, 2003-2004, 2005-2008, and 2009-2010. Items within each temporal section are listed in priority order. Certain items scheduled later may be moved up if the requisite items in the previous section have been completed.

2003-2004:

- Establish Electronic Publications Working Group
- Complete implementation of dynamic online publications catalog
- Establish general member authentication scheme
- Implement password-protected online membership directory
- Draft and implement SAA Privacy Policy for personal information

---

[2] See Appendix X. Cal Lee expresses similar ideas in the parlance of the *Open Archival Information Systems Reference Model.*

- Draft and implement new contributors licensing agreement to enable repurposing and reaggregations of submitted future content
- Investigate third-party credit card transaction vendors and implement electronic credit card payments if possible
- Establish dual format (to support electronic and hardcopy publication) submission standards and workflow for *American Archivist, Archival Outlook* and *Monographs*.
- Create and enable Content Management System taxonomy
- Establish internal standards for archival file formats and facilities for redundant storage of assets (these standards must be regularly updated and maintained)
- Begin discussions with other archival associations re: establishment of archival publishing cooperative to share publication and archiving costs.
- Seek written permissions of past authors to enable reaggregations and web publication of existing content as necessary

- American Archivist:
  - Encode and mount back issues on the SAA website where content is already available in electronic form
  - Issue RFP for retrospective digitization of hardcopy backlog and comprehensive indexing
  - Negotiate licensing contracts with external library vendors for alternative *American Archivist* distribution

- Archival Outlook:
  - Copy "Bulletin Board" and "Professional Opportunities" to the web as discrete sections until web-based AO is available. [Professional Opportunities completed]
  - Increase display advertising from 30% to 50% to increase revenue
  - Post PDF version on SAA Website after three month delay
  - Offer both online and print job advertisements but offer online advertising at a lower price as an incentive to move away from print
  - Survey membership needs for AO content

- Monographs:
  - Investigate member and non-member market for new electronic monographic publications (e.g. Technical Bulletins, Case Studies, White Papers and reports, Electronic auxiliary/supplemental content, Multimedia, Ultra-Basic manuals
  - Select one or two popular monographs in existing electronic format for sample retroconversion project
  - Issue RFP for outsourcing monographic retroconversions to compare against in-house project above
  - Create purchaser licensing agreement to limit redistribution

- Education Products:
  - Enable on the fly content management for workshop schedule and programs [Completed]
  - Link registration form to required readings and/or availability information as copyright permits
  - Continue unrestricted online access to Education catalog
  - Continue online *Directory of Archival Education*
- Annual Meeting:
  - Enable on-the-fly content management for online annual meeting program [Completed]
  - Encourage SAA 2003 and 2004 Annual Meeting speakers to self-post their presentations and link from SAA website (requires licensing agreement)

- o   Develop single shared online registration process for pre-conference workshops and annual meeting

- SAA Section Newsletters:
    - o   Distribute only in electronic form via PDF or HTML
    - o   Survey membership on whether to limit access to SAA members only or make publicly available

- Gray Literature:
    - o   Investigate market for exposure drafts and summaries of research reports
    - o   Investigate Student Chapter interest in soliciting, reviewing, encoding and loading:
        - ▪   Pease Award non-winners
        - ▪   Other student papers judged by submission criteria and neutral review committee

2005-2008:

- Investigate and implement online purchase payments if feasible
- Issue RFP for online author submission interfaces and implement if possible
- Issue RFP for redundant storage and data maintenance services and implement if possible
- Establish and implement metadata standards for submitted content
- *American Archivist*:
    - o   Open available back issues to the public and establish one-year embargo for non-member/non-subscriber access to new editions.
    - o   Secure funding and award contract for retrospective digitization and indexing
    - o   Investigate online review and editing functions for *American Archivist*, issue RFP for software
    - o   Add email discussion area for current *American Archivist* articles

- *Archival Outlook*:
    - o   Simultaneously post electronic version for members only and mail print version
    - o   Embargo public access to AO for three months
    - o   Permanently move "Bulletin Board" and "Professional Opportunities"to the SAA website and remove from AO content
    - o   Post AO/Newsletter Index online in HTML
- Monographs:
    - o   Offer E-delivery of a small number of selected monographs (retroconverted or new) in PDF and CD to test business model
    - o   Issue SAA selected internal reports as web documents free to membership
    - o   Initiate new monographic products if market research and business model permit:
        - ▪   Technical Bulletins
        - ▪   Case Studies
        - ▪   White Papers
        - ▪   Electronic auxiliary/supplemental content
        - ▪   Multimedia
        - ▪   Ultra-Basic manuals
- Annual Meeting:
    - o   Implement SAA hosting for Annual Meeting Presentations and Speaker Biographies as resources permit.
- Gray Literature:
    - o   Implement self-encoding and posting of research reports if desirable
    - o   Implement posting of student papers if student assistance is available

2009-2012:

- Investigate retroconversion of *Archival Outlook*
- Move to broader or comprehensive delivery of electronic monographs as business model permits

APPENDIX I
Preliminary and Advisory Report of *The American Archivist* Subcommittee
Philip Eppard and Nancy Bartlett

Introduction

As the premier archival journal in the United States and one of the leading archival journals in the world, the *American Archivist* plays an important role in the dissemination of current research and thinking on archival subjects and in providing information and commentary on new publications in the field. The semi-annual journal is one of the principal benefits of membership in the Society of American Archivists. As a high-profile print product, it has effectively served both individual readers and the Society membership as a whole. The journal has been an ambassador, of sorts, for the Society and for the American archival community to the rest of the world.

While continuation of the print version of the *American Archivist* is essential, electronic publication of new content, in an electronic version of the journal, would both enhance the journal's well-established role in the archival community and afford new opportunities for diversification of options for the journal's editorial process, production, presentation, and delivery; for alignment in new relationships with other SAA events, programs, services, and products; and for possible aggregations and disaggregations of content according to editorial and membership recommendations and interests. This electronic version of the Society's journal would further validate the Society's interest in and commitment to both new information and new information media. Without introduction of an electronic version of the journal, SAA could lose its established strengths in journal publishing as some authors may turn elsewhere for a creative and timely publication source. SAA might also lose readers who would look for convenient and timely access to authoritative information about archival practice or new research and development in other electronic resources. By contrast, the development of an electronic *American Archivist* would serve to further leverage SAA's assets—its collective expertise and its time-honored commitment to the generous sharing of professional knowledge in reliable sources.

Of equally great value would be the electronic availability of the extensive back file of the journal from its beginning in 1938. The wealth of material in the *AA* would then be easily accessible for use in graduate archival education programs and by researchers and writers on a wide variety of archival topics.

A third feature of electronic publishing for the *American Archivist* is the development of an online review and editing process that would streamline these activities and therefore facilitate a more timely editorial, production, and publication schedule.

Improving Retrospective Access to the Journal

Providing full text versions of more than sixty years of the *American Archivist* is a major undertaking, but it can be done along the lines outlined in the appendix to this report that addresses in detail the retroconversion of the journal. At the same time, access to the contents of the back file of the journal can be improved with the following two steps that are more easily accomplished:

1. Compilation of a complete table of contents for all back issues. This would at the very least allow for browsing of the titles of articles and reviews published since 1938.

2. Compilation of a full retrospective index to the *American Archivist*. The options here would include:

    a.   Collation of existing annual and cumulative indices to create one complete index. (These indices, except for the last few years, do not exist in electronic form.)
    b.   Creation of a simple new author/title index that would at least allow key word searching.
    c.   Creation of a full author/title/subject index.

The first option would be laborious and result in a product that would be uneven because of different indexing practices employed over the years. If the retroconversion plan were pursued as recommended, the second option would be unnecessary since keyword searching of authors and titles would be a feature of the electronic backfile. The third option is the most desirable, as a complete author/title/subject index would overcome the difficulties and inefficiencies in full text keyword searching. The Association of Canadian Archivists has begun an indexing project for *Archivaria* that includes the construction of a thesaurus of archival terminology. The indexing project is scheduled to be completed in June 2003. Conversations with ACA suggest that SAA would be able to share and expand on this thesaurus.

The SAA web site currently has the table of contents for the *American Archivist* from 1997-2001, and abstracts of articles from 1998. For the fall/winter 2001 issue, the full text of book reviews and review essays were made available for the first time, in advance of publication. As a first step to making available electronically the full text of back issues of the journal as soon as possible, texts of those recent issues that still exist in electronic form could be mounted on the web. Meanwhile, a systematic plan for digitizing the entire back file of the journal, probably working backward in five- or ten-year increments should be developed.

JSTOR may be an option for producing electronic versions of back issues of the journal, however JSTOR has refused in the past because they consider the *American Archivist* to be a library and information science journal, not a history journal. We could try to make the case that it could be considered a history journal.

Future Publication of the Journal in Electronic Form

Since the text of the *American Archivist* is currently produced in electronic form, there is no reason why full text could not be made available immediately via the SAA website. As noted above, a small step has been taken in this direction by providing the text of the reviews. The full text of current issues should be made available electronically, with password access for SAA members and journal subscribers. Open access to the full text of the journal should be made available one year after publication.

At the same time, SAA should consider agreeing to let the full text of the *American Archivist* be delivered by bibliographic database services such as ProQuest, and H. W. Wilson. This would be a service that would yield at least a small amount of royalties to the Society. More important, however, is that having the full text of the *American Archivist* available thorough such services could bring the journal to the attention of researchers who might be unaware of it or of its diverse content. The question of preserving the benefits of membership by restricting access to the current issues must be addressed here. Rather than restrict access for a full year, it might be advisable to restrict access only to the current issue. Vendors might be reluctant to mount full texts and pay royalties for journals that are freely available through SAA.

Publication of the *American Archivist* in electronic form could provide an opportunity for enhanced discussion of ideas, issues, and practices discussed in the journals articles. The editor currently receives almost no letters to the editor responding to articles, but the possibility of providing a mechanism for commenting electronically to the articles in the journal as posted on the website should be exploited as a means of stimulating discussion and perhaps inspiring additional articles.

Electronic Editorial Process

Online review and editing will benefit the editorial and production process of the journal in several ways. First, it would speed up the review process by bypassing the postal service. It would enable board members and authors to check on the status of articles at any time. (Currently, some manuscripts are submitted electronically, while others are submitted in electronic form only upon acceptance. All submissions would now need to be in electronic form.) Copyediting of manuscripts can be done onscreen, thereby speeding up the now laborious process of transferring the copy editor's changes on a paper copy to the existing electronic file.

IPC, printer of the *American Archivist*, is using a product now for other publications that could be used by the *American Archivist*. The editor and publications director are scheduling a meeting with IPC in August 2002 to review this product and its potential application to the journal. Moving forward on this initiative will likely happen independent of the rest of the work of the Task Force on Electronic Publishing.

The *American Archivist* and Gray Literature

For many years the *American Archivist* has had a problem in getting submissions. Provision of a vehicle for individuals to post works in progress, project reports, and other gray literature on the SAA website for review and comment could result in additional submissions to the journal, particularly if the capability for readers to make comments enables an author to revise and strengthen a paper in ways that would make it more acceptable for publication. Similarly, the posting of selected student papers emerging from SAA student chapters or on the recommendation of archival educators could result in the identification of potential new sources for articles

Archival Cooperative

One approach to managing electronic publishing of the American Archivist is to share the burden and reap the benefits of cooperation with other professional associations. The History Cooperative oversees electronic publication of the *American Historical Review*, the *Journal of American History*, the *William and Mary Quarterly*, and five other history journals. We recommend that SAA initiate conversations with the Midwest Archives Conference, the Association of Canadian Archivists, the International Council on Archives, Society of Georgia Archivists, and any other professional archival association that publishes a journal to investigate the possibility of forming an Archival Cooperative. The possibility of collaborating with ACA in updating and maintaining a controlled vocabulary of archival terminology for indexing suggests that there are opportunities for collaboration in this area. As the leading society in the field, it is appropriate that SAA take the lead in this area. The Society could seek funding from the NHPRC for such an endeavor, although that would make inclusion of ACA problematic. A fuller understanding of the business model of the History Cooperative is needed and inquiries to other organizations should be made to determine the level of interest. If SAA were to try to seek external funding for moving the journal into electronic format, collaboration with other organizations would likely strengthen our chances.

Business Models

- Full text of *American Archivist* backfile available for free public access through SAA website.

- Current issue and immediate prior issue accessible only to SAA members and journal subscribers to the print version.

- Full text available through vendors on a royalty basis. Current issue embargoed for six months.

- Develop RFP for full subject indexing of entire backfile.

- Develop Archival Collaborative to share burdens of ongoing maintenance of electronic texts and indices.

APPENDIX II
Archival Outlook: The Next Generation
by Teresa Brinati, Johanne Pelletier

Value Statement

The most recent membership survey indicates that *Archival Outlook*, the bimonthly newsletter of the Society of American Archivists, is considered the most useful benefit of membership in SAA and that there is a strong demand by the membership to see this service continued. This and other anecdotal evidence suggest that the newsletter has an intrinsic value due to its content (a record of the activities of the membership, the association, and the profession at large); its unique graphic image; and its portability. In addition, *Archival Outlook* also has financial value to the organization through the revenue stream generated by display and job advertisements.

In summary, *Archival Outlook* in its current print form offers the following to SAA:

1) Intrinsic value;
2) Financial value through revenue generated by advertisements;
3) Record value—where the newsletter acts as a critical resource on the history of the organization and the community it represents.

Vision for Archival Outlook

Print: *Archival Outlook* should continue to be a benefit of SAA membership and as such should continue to be published in print form. Component pieces better suited to electronic format, such as "Bulletin Board" (calendar and funding listings) and "Professional Opportunities" (job ads) should be removed from the newsletter for exclusive dissemination and access on the SAA website.

The following also should be considered: The fee structure for "Professional Opportunities" should be revisited and readjusted so that any advertiser wanting to run a job ad online *and* in the newsletter would be required to pay a much higher fee, which would serve as a disincentive to do so. After the Bulletin Board and Professional Opportunities are removed from the print piece, the newsletter should eventually be reconceptualized.
In order to do this, the membership would be surveyed to ascertain their preferences with respect to the content of the newsletter. The ultimate goal for the print piece is to elevate the image and stature of SAA through a more sophisticated bimonthly magazine. Substantial feature articles—such as SAA's coverage of Executive Order 13233, the events of 9/11, and a critical review of *Double Fold*—benefit the membership by providing in-depth analysis on important professional matters and also place archival work in a greater societal context. The greatest challenge facing this endeavor would be cultivating a bank of contributors, identifying appropriate subject matter, developing an editorial calendar, and establishing an expedient review process that utilizes the latest technology. To defray printing expenses, display advertising would be increased from 30% to 50% of the publication. In addition, the page length of each issue would be limited to contain costs.

Electronic: *Archival Outlook* should be available in PDF on the SAA website:
1) Issues should be posted online at the same time the print version is sent to press;
2) Issues should be posted in a "members only" area on the SAA website with access limited to membership for the first three months of publication;
3) After the three-month embargo, issues should be accessible to the public.
4) The *Archival Outlook/SAA Newsletter Index, 1975-1995* should be posted online in html. The index should be updated and maintained annually.

While component pieces of *Archival Outlook* may be available in other "news" sections of the SAA website, the full intrinsic value of the newsletter as a packaged product for members would be protected for the membership for a three-month period.

Retroconversion:

Retroconversion of back issues of *Archival Outlook* to PDF is not a priority. Timely dissemination of new issues is the main concern.

APPENDIX III
Report of Monographs Subcommittee of SAA-EPTF
[David Haury and Dennis Meissner]

Existing SAA Monographs

Definition

SAA's catalog of professional resources lists in the neighborhood of fifty "monographs" published by SAA. Monograph is defined broadly to include volumes in the archival fundamentals series and basic manuals series and smaller works such as case studies and even brochures. Basically for the purposes of this report a monograph is considered anything other than *American Archivist*, *Archival Outlook*, and other newsletters.

Recommendation

Electronic versions of all future monographs published by SAA should be available for distribution through the World Wide Web. All SAA monographs are now prepared for printing using desktop publishing software, and conversion from this software into PDF or another format suitable for distribution over the web should be a simple and low cost procedure. Future changes in desktop publishing software or production methods should place a high priority on further simplification of electronic distribution. This conversion to PDF should also be considered on a case-by-case basis for other recently published SAA materials already available in an electronic version. Some of the older monographs available from SAA are not available in an electronic version, and conversion of these is not recommended until new editions are prepared.

Future monographs published by SAA should take into consideration the advantages of electronic distribution in their formatting and structure. It is anticipated that at some point the version distributed electronically may become the primary version and the paper version will be secondary and eventually eliminated. (Many SAA manuals are already printed on 8.5x11 inch pages and this should become standard for the convenience of those who download volumes and print their own paper copies).

Economic Model or Business Plan

Ideally the SAA catalog of professional resources would be available online and linked to a shopping cart program. When a customer clicks on a publication with both an electronic version and a paper version, they would be offered the option of purchasing and downloading the electronic version at a discounted price and would avoid the shipping charge as well. [TFEP members noted that teaser excerpts might also be presented]. For example, a fundamentals series volume that costs $30 plus about $7 shipping might be offered for a total of $20. Some pamphlets or brochures might be available online for no charge.

Purchasers of electronic versions would be required to agree to a licensing agreement authorizing them to print one copy for personal use and load the electronic copy on one computer (and save one backup copy to removable media).

For those without adequate Internet access for downloading copies to be reasonable, SAA should investigate burning the electronic version of some publications to CD-ROM or other media. This method of distribution could be investigated on a trial basis for a limited number of publications or perhaps certain publications could be packaged together on a CD. The CD version of a publication would be priced less than a paper copy since printing and mailing costs would be lower if the volume of orders is sufficient. This could be an intermediate step toward eliminating paper copies of some publications. In the example above, the fundamentals volume on CD might cost $25 plus $5 shipping.

This model is highly dependent on being able to offer e-commerce and download capabilities on the web. As SAA considers options for online payments for memberships, meeting and workshop registrations, and other products or services, the capacities and design necessary for the online purchase of publications should be a high priority. Online payment for membership renewals and meeting registrations would be a significant convenience for many members but would probably not result in a measurable increase in

memberships or registrations or reduce the costs of those transactions. However, online payment for and delivery of publications would be very likely both to increase orders and to improve profit margins. Thus the long-term plan for e-commerce should include capabilities for online ordering and delivery of monographs.

In a different business model, SAA might offer the available electronic versions of some monographs free of charge on a member's only section of the SAA web site, or perhaps all publications would be available for an annual subscription. However, it is assumed that most publications would be available online to non-members and members with members receiving a discount.

Opportunities For New Monographic Formats

One objective of the "monographs" subcommittee of the EPTF is to suggest, and develop business models for, new types of publications that have not been attempted in the paper-based environment, but which might be both possible and sustainable in an electronic environment. What follows is a preliminary attempt to suggest some publication types that might fit these criteria, and to rough out some initial elements of a business model for each.

Technical Bulletins

SAA has never published the sort of short, tightly focused bulletins and leaflets that convey very specific and highly practical information about a process or technique that archivists need to know. Yet many related professions (like the preservation community and AASLH) do so all the time, and find a sustainable market for an ongoing stream of such publications, often issued in one or more numbered series. We, to the contrary, have bundled such information into book-length manuals, etc. The disadvantage with that bundling model is that in a rapidly evolving information environment, some information ages faster than other information. It might make more sense to disaggregate such content from its current carrier, and permit the user to purchase it cafeteria-style, or to aggregate content post facto.

These leaflets, as PDF files, might range from 3 to 20 pages, and may be priced according to size, according to the labor inputs that went into their creation, or according to some other formula. Content would be extremely focused: how to surface-clean parchment; authenticating handwritten 19th century U.S. government issuances; creating a valid XML instance; preserving shellac-based audio discs produced during the 1910-1925 period. You get the picture. Acquisition fees for each title might be small to none, since they would represent valuable, but not arduous publishing opportunities for their expert authors. Because of the small page run for each title, they will serve out very well as PDF files, and the resulting paper output can be bundled into binders, etc. Each title can bear distinctive branding to indicate its place in a particular SAA publication series.

In summary, ePublishing such series would lower costs (making it conceivable to publish something we don't publish at all), permit the user to aggregate content in any desired manner, would lend itself to electronic dissemination, and might expand the SAA customer base by creating a product with appeal to less professionalized communities, like county and local historical societies, public libraries, etc.

Case Studies

SAA already publishes the 8 case studies in electronic records. It is very possible that this is an expanding market that will be driven by the needs of Archival Studies courses. Presumably, this market could extend to all the functional areas and processes that concern archivists. Electronic creation and distribution could greatly lower costs and still market the titles in a convenient form, in the same way that might be the case with technical bulletins. So, the business model considerations noted in that situation would also apply fairly well here. The main business decision would be to determine demand levels.

White Papers and Reports

SAA, its sub-entities, and other bodies within the archival community periodically issue vision statements, political position pieces, and the results of significant professional studies and investigations that are of

interest to the membership, to subsets of it, and to the larger society. Given the expense of producing and delivering such materials, SAA becomes loath to do so except for the most important issuances. ePublishing could change those economics, and the website then provides a vehicle for organizing this content effectively, and for assuring its continuing accuracy and freshness. Business decisions will center on whether to give away all such content, or to divide it into several categories and use a different merchandising model for each.

Auxiliary Content for Printed Monographs

The Pubs. Board has, I believe, in the past run across monographs that might have been more robust titles if they could have incorporated content better suited to ePublishing. Such content might be an electronic final chapter or an appendix that contained either dynamic or multimedia information that did not lend itself to expression in a printed form. Cal Lee notes three aspects of dynamic content: Frequently updated or reissued; Separate digital objects that may be assembled on the fly; Content that is constantly changing in response to user activities or profiles, such as a portal. SAA e-monographs may be frequently updated. We might create better print monographs if we had the option to store and disseminate some of their content electronically. An example here is the X*ML Bible*, a printed work one of whose chapters (dealing with XSL style sheets) is found online because of its frequently updated content.

By producing such hybrid monographs we can deliver more robust content as well as making it possible to consider titles that we might have been forced to reject before. The main business consideration is how to "sell" the additional content. It could simply be made freely available to all visitors, perhaps along with the monograph's TOC (this is the *XML Bible* model); could be offered free to all members; could be offered cheap to all members (if the content were valuable enough); or could be offered only to purchasers of the printed work (is this technologically feasible as a normal business practice?).

Monographs With Critical Non-textual or Multimedia Content

Monographs shot through with critical non-textual content could be published entirely in the electronic environment. Presumably, because of the nature of the content, no PDF or other print surrogate would be offered. Business problems include developing a model for selling a completely virtual "book," which includes questions of pricing, delivery, buyer authentication, and persistent accessibility to the buyer.

Basic Manuals

Perhaps ePublishing will make it attractive to create an "ultra-basic" set of manuals for very green archivists and, more likely, for allied communities who may not want to purchase or read book-length manuals. Not at all sure of the demand for such a product, but if price is the deciding factor, ePublishing might offer the solution. Content could be fairly cheaply derived from existing manuals.

Business Models for Suggested New Monographic Formats

The business models outlined below all assume that the monographs are identifiable and selectable via the SAA website; that some sort of electronic cash register or ordering mechanism exists; and that the web pages storing ePublications can be partitioned to accommodate varying levels of access.

Technical Bulletins

- Cafeteria-style selection (via shopping cart) from individual bulletins grouped into thematic series. Alternately, an entire series could be aggregated on CD.

- All titles share highly recognizable formatting and branding

- Stored and delivered as PDF files

- Individually priced based on (1) acquisition costs, (2) storage/delivery costs, (3) demand

- Content updated as needed: titles individually updated and re-published as content ages

Case Studies

- Cafeteria-style selection (via shopping cart) from individual titles grouped into thematic series. Alternately, an entire series could be aggregated on CD.

- All titles share highly recognizable formatting and branding

- Stored and delivered as PDF files

- Individually priced based on (1) acquisition costs, (2) storage/delivery costs, (3) demand

- Static content: title pulled when content ages

White Papers and Reports

- Cafeteria style selection (via shopping cart) from set of unique titles

- Titles share some minimal branding indicating their genre

- Stored and delivered as PDF files (barring excessive length)

- Individually priced based on (1) acquisition costs, (2) storage/delivery costs, (3) demand

- Some titles may be offered free to all or free to members

- Static content: title pulled when content ages

Ultra-Basic Manuals

- Cafeteria style selection (via shopping cart) from set of unique titles

- All titles share highly recognizable formatting and branding

- Stored and delivered as PDF files

- Marketed to members of allied professions, semi-professional archivists, and amateurs

- All titles identically priced, based largely on demand and marketing costs

- Content updated as needed: titles individually updated and re-published as content ages

eBooks

This refers to book-length monographs that must be ePublished in their entirety because they are rife with non-textual content that cannot be presented in a print format.

- Titles are unique and not necessarily grouped with other eTitles

- Stored as set of linked files, probably in multiple formats, that can be served out to purchaser as a seamless whole with a book-like appearance

- Delivered either as a palpable object (CD) or as a persistent resource (mounted on SAA server) accessible via password

- Individually priced based largely on significant acquisition and storage/delivery costs

- Content frequently updated, with all individual file components amenable to frequent updates (this may argue against distribution on fixed medium like CD)

Books with Electronic Addenda

- Titles are unique and are most likely part of the traditional monographs catalog

- Books are delivered by traditional means; electronic addenda may be delivered by (1) CD accompanying book or as a persistent resource (mounted on SAA server) accessible via password

- Priced using traditional monographs model, modified by additional costs involved in future updates to and storage of electronic addenda.

- Content of book is static; content of ePublished addenda is updated as needed

APPENDIX IV
Report of Education Subcommittee of the SAA-EPTF
By Solveig De Sutter and Patti O'Hara

The mission of the Society of American Archivists is to serve the educational and informational needs of our members and to provide leadership to ensure the identification, preservation, and use of the nation's historic record.

The Education mission is to develop and offer quality continuing education that is relevant and accessible to professional archivists via a variety of delivery methods.

Existing Education Products:

Workshops and Seminars
The SAA Education Department works with regional and local co sponsors to provide as many as 30 live, remote, workshops and seminars annually, including 6-9 SAA pre conference offerings. The length of the different offerings varies from 1-3 days averaging a total of 43 days of SAA CE/year throughout the US. In addition, we are working with an eLearning service provider to establish a foundation for providing regular, reliable on line learning opportunities.

Depending on the workshop or seminar, pre course readings, workbooks, handouts and publications are distributed with the registration materials either by mail in advance or shipped to the site for on site distribution. Currently, only one of our instructors provides the registrants links to pre course readings available on the web.

Finally, the registration process-on line and hard copy- requires registrants to provide professional and membership information which is then duplicated by Education and forwarded to the instructors for developing appropriate learning objectives.

Education Catalogue of Workshops and Seminars
The targeted audience for this online document is prospective co sponsors to help them match their needs to the offerings. This is in static form on the SAA web site and is currently out of date.

Schedule of Upcoming Workshops
The target audience for the schedule of Education workshops and seminars is prospective member and non-member registrants. The on line schedule includes links to a secured on line registration form as well as a PDF form for mailing or fax. The schedule and registration form need constant revisions as new events are added or modified for various reasons.

Needs Assessments and Program/Instructor Evaluations
The last formal education specific needs assessment was compiled March 1997. Since then, planning and course selection has been based primarily on specific needs assessed and expressed by regional and local archival groups.

Workshop and seminar evaluations are distributed and completed following each event. The results are manually tallied and compiled for evaluation and assessment purposes.

Directory of Archival Education
Finally, the Directory of Archival Education has been on line since July 2001. The target audience is current and prospective students in archival studies. Since November 2001, Education providers have been able to manage the content of their site contribution as needed. The printed version has been discontinued. The Education office will download and send the on line directory upon request.

Recommendations:

Workshops and Seminars
Continue to provide live, remote, workshops and seminars but gradually introduce more and more on line offerings. The transition to increase on line offerings could be eased by making more and more of the pre workshop/seminar readings available on line. This would also expedite the workshop material delivery process and cut mailing and duplicating costs. In addition, pre workshop on line introductions or learning objective inventories initiated by the instructors or the Education Directors could replace the need for registrants to provide professional and membership information which is currently duplicated by Education and forwarded to the instructors.

Completing a registration form should be linked to a list of required or recommended readings specific to that selection and information on how to purchase them.

Education Catalogue
Continue to provide the catalogue exclusively on line and with unrestricted access. Consider data mining for internal market research

Schedule of Upcoming Workshops
We look forward to the capability of managing the content of the workshop logistics and having the changes automatically made on the registration form.

Directory of Archival Education
Continue to offer the Directory of Archival Education on line and the Education office will download and send the on line directory upon request. Continue to assess the education providers' needs regarding the content, input process and display of data.

APPENDIX V
Report of Annual Meeting Program/Presentations Subcommittee of SAA-EPTF
Solveig DeSutter, Patti O'Hara, Scott Schwartz

Existing Annual Meeting Program/Presentations

The annual meeting is held in late summer in different cities throughout the country and includes a wide array of informative program sessions, pre-conference workshops, tours of local repositories, special events, exhibits and networking opportunities. The objectives of the annual conference are the promotion of archival education, communication of information, promotion of fellowship and professional networking, and the management of the business of the organization. Pre-conference workshops are frequently associated with the annual meetings, but the registration and management of the educational workshops and annual conferences are administered separately. While there is a distinction between the administrations of these two programs, it is recognized the annual meeting provides a valuable marketing tool for the educational workshops. This is in addition to advertising of workshops through archival listservs, press releases, and targeted mailings.

The traditional 8x11-inch glossy preliminary program that includes historically unique photographs and is printed for each annual conference is SAA's most important traditional marketing tool. This program is important to members because it provides a uniquely visual, tactile, and informational context that entices them to browse the program's content (i.e., the warm and fuzzy element of publication). In addition the preliminary conference program is promoted as a membership benefit that is delivered exclusively to members of the organization to help them plan their participation in the conference activities. The current on-line version of the preliminary program that is available to everyone lacks the "added value" promised by on-line technologies and is less visually and intellectually engaging. A final conference program is provided to all registered participants that contains the most current information about the program and supersedes any information contained in the pre-conference program. This final printed program is less glossy and functions as a type of navigation tool to guide participants for sessions of specific interest to them.

Conference presentations are selectively recorded and cassette audiotapes are provided to participants for a fee. No handouts or other visual documentation used during the original presentations are included with these sound recordings.

Definition

The annual conference provides a means for effective cooperation among persons concerned with the documentation of human experience, and serves to stimulate research in archival administration and records management; to promote the adoption of sound principles and standards by all public and private agencies responsible for the preservation and administration of records; to foster a better public understanding of the nature and value of archival operations and holdings; to develop professional standards, particularly for the training of archivists, records managers, and custodians of private papers, and to improve the facilities and increase the opportunities for such training; to maintain and strengthen relations with historians, librarians, educators, public administrators, and others in allied disciplines; to cooperate with other professional organizations, cultural and educational institutions, and international organizations having mutual interests in the preservation and use of recorded heritage.

Recommendation

1.   Develop a single shared automated registration process for both conference workshops and the annual conference meeting so registrants need only be registered once.

2.   Develop on-the-fly content management for the annual meeting program to keep the on-line version of the annual conference program current. In addition provide real-time e-

mail updates of program changes to all individuals registered for the annual conference.

3.   Utilize the sound bytes of select conference and workshop sessions, recorded before the conference, to be added to the on-line program to serve as one interactive marketing tool.

4.   Do not eliminate the glossy printed preliminary conference and workshop program.

APPENDIX VI
Report of Section Newsletters Subcommittee of SAA-EPTF
[Scott Schwartz]

Existing SAA Section Newsletter

As a service to SAA Sections, the SAA office prints and distributes their newsletters or other mailings up to three mailings each year, with a total number of pages for the year not to exceed eighteen (18) single-sided sheets of camera-ready material on paper measuring 8 _ by 11 inches.   Deadlines for newsletter submissions are December 1, March 1 and June 1. Section chairs and editors are asked to inform the SAA office at least two weeks prior to their submission so that mailing labels may be processed. The SAA office requests that no newsletters be submitted between July 31 and September 7.   Section chairs and editors are responsible for the content, accuracy, and legibility of the newsletters and other mailings. The SAA office ensures that mailings are duplicated, labeled, and mailed. All mailings of more than 200 items are sent bulk rate resulting in a deliver time of three to four weeks. All mailings from Section leadership to their membership are also distributed to Council and the Executive Office. SAA Council includes as a line item in the annual budget an amount sufficient to provide each section with three mailings to its membership during each year.

Definition

Sections focus attention on and give voice to the concerns that form the Section's particular areas of interest. Sections focus expertise and interest in the professional functions and responsibilities to carry out the archival profession's mission to identify, administer, and promote the use of records of enduring value. Section newsletters provide a forum for members to express matters of interest within the Society, and provide channels of communication from the Sections to   the membership of the Society and to Council on matters of concern to the Section.

Recommendation

All section newsletters should be distributed only in electronic form. This electronic format may be either a PDF or web-based document that is made available to section members through SAA's website. SAA no longer should duplicate and distribute paper versions of section newsletters.

The increasing cost of duplication and mailing of section newsletters makes the continuation of the current publication and distribution of these newsletters incapable with SAA's current budget constraints. The restrictive time frame for newsletter submissions and the greater dependence on bulk mail services makes the timely communication of information between members of sections and a newsletter that serves as a voice of Section's particular concerns to SAA's general membership less viable. Electronic newsletters will provide SAA's entire membership with greater timely access to information and concerns related to the various Sections' specific areas of interest. Electronic newsletters will also reduce the cost to SAA that is currently associated with the duplication and distribution of paper documents.

APPENDIX VII
Report of the TFEP Subcommittee on Gray Literature
Rob Spindler

TFEP members discussed the possibility of enabling electronic publication of several types of gray literature on the SAA website. Gray literature could include:

- Research Works in Progress

- Student Papers

- SAA Annual Meeting Papers

The advantages of gray literature include speed to publication, easily accessible opportunities for student and early career professional papers, enhancing the breadth of archival research, and providing opportunities for review and discussion of new or emerging ideas. Availability of gray literature and effective metadata could also attract new visitors to the SAA site.

Research papers in progress may be hosted at the SAA site, although policies and a process should be established to control the amount of content SAA must host. Disclaimer language identifying this material as unedited or unreviewed and denying any endorsement of the content is recommended, and a release form would be required. Authors should consider whether other potential publishers would view submission of these works as prior publication.

Student Papers are a rich source of archival research that could be published by SAA. TFEP members discussed extending the functions of the Pease Award to include posting of non-winning submissions. A separate flow for student content could be established with papers submitted from students at large to a committee of representatives from student chapters. In this scenario the students should solicit, review, select and format for publication a limited number of titles each year. A member of the SAA Publications Board or a volunteer coordinator might facilitate this activity.

TFEP generally supported the idea of posting Annual Meeting presentations, although the source of labor for reformatting the content and the quality of the presentation medium were recognized as important challenges. There were differences of opinion on whether a PowerPoint presentation had sufficient content and context to warrant publication. PowerPoint provides opportunities to display some graphical and visual elements that audio files cannot deliver, but also presents some challenges in terms of preservation and delivery file size. Textual word processing files might enable more complete texts to be mounted, but incorporating visuals was seen as a problem. Submission standards and a release form would need to be established to move on this item.

APPENDIX VIII
The Retroconversion of *The American Archivist* to Digital Format
Ed Galloway, University of Pittsburgh
Member of the SAA Task Force on Electronic Publishing

This report, accompanied by my recommendations, will briefly describe the issues and challenges for converting the physical volumes of *The American Archivist* into digital format for searching and viewing on the Web. Without evidence from SAA members to support limited retroconversion (e.g., only past 10 years), this report will assume that the complete retroconversion of *The American Archivist* is warranted and desirable.

I. Acquisition of complete journal set

In order to embark on a retroconversion project, SAA must be able to acquire a complete set of *The American Archivist* for digitization. Ideally, this set should be able to be disbound for scanning to decrease the cost per page capture and facilitate ease of capture. The SAA archive maintains two complete sets of *The American Archivist*. Teresa Brinati, SAA director of publishing, also maintains one complete set of *The American Archivist*.

Recommendation
- Take one complete set from the SAA archive for retroconversion; or
- Ask members to donate volumes in hopes that a complete set can be acquired.

II. The physical volumes

The American Archivist has been published by SAA since 1938, and has remained at a constant trim size of 6.75" x 10". To date, SAA has published 64 volumes (248 issues) totaling 30,805 pages. From 1938-1997, each volume contained four issues; from 1998 to the present, each volume has only contained two issues. The volumes are composed of printed black text on white paper. While the issues frequently contain tables and charts, very few contain half-tone illustrations. Volume 63 (1) is the only issue to contain full-color plates (16 of them). The covers of the journals contain a mixture of halftones, continuous tones, and mixed.

A brief study of *The American Archivist* from 1991 to the present revealed the following type of information included in the journal. This list does not indicate when certain types were introduced or eliminated.

- Cover (front/back)
- Boiler plate
- Table of contents
- Gallery of contributors
- Forum
- From the editor
- Articles
    - Title
    - Author(s)
    - Abstract
    - Tables/Figures
    - Footnotes
- Perspectives
- International Scene
- Reviews
- Review essays
- Society of American Archivists

- o Executive Director's annual report
  - o Treasurer's annual report
  - o SAA Council meeting minutes
- Editorial policy
- Advertisements (paid/house)
- Errata
- Annual index

It is important to note and capture this structure to implement full-text searching by category, and/or offer the ability to navigate the digital versions via a virtual table of contents.

Recommendation
- Perform a detailed structure analysis of the complete set of the journal to determine:
  - o What characteristics of the journal have changed over the years?
  - o What components have appeared, disappeared, and reappeared over the years?

III. Possibilities for delivery of a digital version

There are at least four possible methods for delivering past issues of *The American Archivist* electronically: GIF/JPEG images of each page, PDF (individual articles or entire issues), ASCII, or HTML (or a combination of any of these).

Regardless of the format to be delivered to the user, a retroconversion project of this nature must assume that the individual pages of the journal will be scanned, including the covers. The scanned images will be used as surrogates for the physical volumes, and/or for extracting the text for searching or display as ASCII or HTML.

Recommendation
- Create a set of master (archival) images for long-term preservation.
- Create lower resolution derivative images from the master images for Web access.

The master image files should be scanned as 600 dpi, 1-bit TIFF 6.0 images using ITU-G4 compression.[3] Although *The American Archivist* does contain half-tone illustrations, capturing this information bitonally will not compromise its viewability. However, the 16 color pages from Volume 63 (1) and the journal covers should be scanned as 400 dpi, 24-bit color to capture the full-range of colors.

TIFF images can be converted to GIF images on-the-fly for Web use. TIFF images could also (or instead) be converted to PDF. However, the ability to read PDF files places an additional requirement on the user to download Adobe Acrobat's free reader. Either of these scenarios will not enable full-text retrieval because texts will be represented only by bit-mapped images.

Recommendation
- Utilize "just-in-time" delivery of images by converting TIFF images to GIF on-the-fly.[4]
- For printing purposes, incorporate a feature enabling users to convert an image or entire article to PDF.
- Store copies of the master page images (TIFF) on the host website.

---

[3] Greenstein, Daniel and Gerald George, "Digital Reproduction Quality: Benchmark Recommendations," RLG DigiNews: Volume 5, Number 4 (August 2001). http://www.rlg.org/preserv/diginews/diginews5-4.html#featured
[4] For examples of this technology, see the Making of America website (http://moa.umdl.umich.edu/) or the Historic Pittsburgh website (http://digital.library.pitt.edu/pittsburgh/).

IV. Browsing the digital version

Users should have the ability to browse the contents of *The American Archivist* by volume and issue number. A user should have the option to select a specific issue and browse an online version of the table of contents. Once a user can navigate the table of contents for each issue, he/she should be able to select a link to any section included in that particular issue. Once a user clicks on a link to read a specific article, he/she should have the ability to turn pages forward and backward as well as be able to jump to any specific page in the issue.

V. Searching the digital version

Utilizing optical character recognition (OCR) to capture the text information from the page images will enable full-text retrieval and/or ASCII/HTML viewing. Since the source material is typeset, an OCR engine should have a high confidence when interpreting the source. Nonetheless, a decision will need to be made regarding the acceptable level of accuracy in the resulting files. 'Dirty' OCR (i.e., the raw output of the OCR process; text which is not proofed for corrections) should be good enough for retrospective search and retrieval in *The American Archivist*. This text, whether it is cleaned up or 'dirty,' can be made available to users or remain behind the scenes.

Recommendation
- Run an OCR engine (or OCR package like PrimeRecognition) to produce text; do not proof for misspellings or re-key.
- Only deliver images of past issues to the user; do not display ASCII to user.
- Do not produce an HTML version of the journal from the text since this would require the labor-intensive process of proofing and encoding the text.

In order to provide users with the ability to search the full-text of the journals, a decision must be made regarding the granularity of search and retrieval that is feasible and/or desirable. For example, *The American Archivists* could be searched by its full-text with no delimitations, or it could be searched by abstract, author, title, subject, or article type (e.g., article, review essays, reviews, Pease award, etc.). An analysis of the structure of the text may indicate additional categories.

Recommendation
- Search full-text by abstract, author, title, subject, or article type.

In order to give users the ability to browse a virtual table of contents and/or search the full-text by categories, various kinds of metadata for the journal must be captured and an SGML engine employed to index it. This metadata includes at least: pagination, type of article, title, author, abstract, and the presence of illustrations or figures.

VI. Metadata to be created

The following metadata should be captured when converting journal issues to digital format (courtesy Cornell digital imaging tutorial[5]):

Administrative metadata

---

[5] http://www.library.cornell.edu/preservation/tutorial/metadata/metadata-02.html

- Use appropriate TIFF 6.0 file headers to record technical information, e.g., ImageWidth, ImageLength, Compression, StripOffsets, RowPerStrip, StripByteCounts, Xresolution, Yresolution, Resolution Unit, and BitsPerSample.

Structural metadata
- Assign file names and directory structures to the image files and the associated metadata files.
- Create a database to store and manage bibliographic information from the cumulative journal indexes to enable structured vocabulary search (e.g., journal volume, issue, title, author name, beginning and ending page number).
- Use TEI Lite SGML encoding to map the basic structural elements of the journals, such as volume, issue, title, author name, abstract, parts, beginning and ending pages for each article, to facilitate online searching and browsing.

Descriptive metadata
- Create Dublin Core records, and/or Open Archives Initiative (OAI) records, and/or Resource Description Framework (RDF) records.
- Assign persistent, globally unique, and location-independent file names.
- Use appropriate TIFF 6.0 file headers for image description to record descriptive elements essential for identifying the file (e.g., project ID, institution, collection, year of publication, title, author, image sequence number).
- OCR images to provide free-text keyword access.
- Create HTML tags with Dublin Core information to facilitate resource discovery.
- Register the Website with relevant subject directories, specialized subject portals, and gateways to increase coverage by Web search engines.

VII. Creating the digital version

SAA will need to determine who should perform the digitization, metadata entry, and quality control steps associated with this conversion project. The quality control steps include proofing the journal images, creating and proofing the structural, administrative, and descriptive metadata, and performing the OCR. It will also be important to select an OCR engine sufficient for the job.

Recommendation
- For a relatively small, one-time project of this nature, it will be in SAA's best interest to outsource the project. A commercial vendor or non-profit institution with a scanning lab can offer the services required for digital conversion, encoding, and quality control.
- Develop an RFP to address the specific needs, criteria and requirements of the project.[6]

VIII. Data management system

SAA will need to determine what kind of data management system should be employed to store, index, search, and deliver data to the user. A data management system can search for and retrieve images; provide an access interface; provide a structural framework for the digital images; track and control source material and generated images during the digitization process; control access to images; automate and validate data entry; facilitate use of controlled language; and provide a home for the metadata needed or long-term access to the images. Such systems include in-house systems using common desktop databases; in-house systems using client/server architecture; specialized image management programs designed to run on the desktop or on a client/server architecture; and SGML/XML-based solutions.[7]

---

[6] The RLG Model RFP serves as an example of how Cornell University adapted the RLG Guidelines for Creating a Request For Proposal for use in a text-based digitization project. (http://www.rlg.org/preserv/RLGModelRFP.pdf)

[7] Hirtle, Peter. "Image Management Systems and Web Delivery." In *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, Anne R. Kenney and Oya Y. Rieger, eds., 2000, pp. 119-134.

If SAA is not in a position to develop or purchase a data management system, it should consider partnering with a commercial organization or non-profit institution that can provide this service.

Recommendation
- Solicit bids from commercial organizations/non-profit institutions to host, serve and maintain the data files and search engine.
- Seek an SGML/XML-based solution.

IX. Interoperability of the data

SAA has the option to make the digital version of *The American Archivist* interoperable with other digital library systems. One example is the reference-linking model (e.g., CrossRef) that would enable article citations to be directly linked to an online version of that article (if one exists). Another interoperable model to consider is the Open Archives Initiative (OAI) protocol for the harvesting of metadata by a third-party to allow resource discovery. One more model to consider is the Resource Description Framework (RDF), an XML-based application to provide a flexible architecture for managing diverse metadata in the networked environment.

Recommendation
- Incorporate OAI protocol into metadata scheme.
- Investigate the benefits of applying RDF metadata to this project.
- Do not incorporate reference linking because this model is too labor-intensive and unproven for incorporating valid links in a document.

X. Ongoing accessibility to the digital version

Once a decision is made about who will host the images/index, SAA must be given assurance that the information will be accessible to the user community and continue to be migrated/refreshed as technology changes.

If SAA wants to provide access to the online version of *The American Archivist* to only its members, then it must employ an authentication step to prevent unauthorized use.
This report assumes that only existing issues of *The American Archivist* will be retrospectively converted into digital format. The system developed for the electronic delivery of the journal must be able to incorporate future issues of the journal in their native electronic form. Therefore, all issues of *The American Archivist* from 1938 to 2001 would be available as scanned images, but 2002 and ongoing versions should be made available as HTML or searchable PDF. These files will need to be integrated and capable of being processed by the data management system.

When delivering full-text content to users on the Web, SAA should adhere to the Uniform Resource Name (URN) for persistently naming and locating digital objects (e.g., issues and individuals articles within *The American Archivist*). The URN is intended to serve as persistent, location-independent, resource identifiers rather than pointing to a location on a server (e.g., URL).

XI. Indexing *The American Archivist*

The paper version of *The American Archivist* is currently indexed at least by:
- OCLS FirstSearch: Article First
- WebSPIRS (SilverPlatter): Title, Author, Source, Abstract
- EBSCO: Current Citations (from Volume 55 (4) to the present)

Each of these index and abstracting services offer their index online via licenses. Regardless of indexing the electronic version of the past issues, *The American Archivist* will continue to be indexed by the above

services. It appears that *The American Archivist* articles, review essays, and review are indexed by the above services.

XII. Funding to support the retroconversion project

SAA should seek non-member support to pay for the retroconversion of *The American Archivist* into electronic format. It is recommended that SAA contact the following organizations, foundations and government agencies that have demonstrated support for projects that employ technology:

- Atlantic Philanthropies (http://www.atlanticphilanthropies.org/index.asp)
- Documentation Abstracts, Inc.
- Institute of Museum and Library Services (http://www.imls.gov/)
- Library of Congress (http://www.loc.gov)
- The Henry Luce Foundation (http://www.hluce.org/)
- The Andrew W. Mellon Foundation (http://www.mellon.org/)
- Alfred P. Sloan Foundation (http://www.sloan.org/)
- The H. W. Wilson Foundation (http://www.hwwilson.com/)
- Robert W. Woodruff Foundation (http://www.woodruff.org/)
- National Endowment for the Humanities (http://www.neh.gov/)
- National Endowment for the Arts (http://www.nea.gov/)
- Pew Charitable Trusts (http://www.pewtrusts.com)
- Ford Foundation (http://www.fordfound.org/)
- Spencer Foundation (http://www.spencer.org/)
- Rockefeller Foundation (http://www.rockfound.org/)

XIII. Who should be able to access the electronic version of *The American Archivist*?

Upon the retroconversion of the past issues of *The American Archivist* into digital format, it is recommended that SAA provide free and open access to users of the Web for several reasons. If funding is secured from any of the above foundations or government agencies, they will almost certainly require free and open access to the material regardless of membership. Furthermore, offering free, electronic access to past issues of *The American Archivist* will enable SAA to promote its services by sharing its body of professional literature with the Web community. It is also reasonable to assume that authors, whose articles were published in *The American Archivist*, would want the greatest readership and distribution of their ideas and work. Moreover, SAA is not likely to recover many costs from non-SAA members accessing this content on the Web.

In the best interests of SAA members, however, SAA should restrict electronic access to the current issue of *The American Archivist*. An embargo period of one to two issues for non-member access seems reasonable.

XIV. Estimated Retroconversion Costs (added to this Appendix November 15, 2002)

 This document will briefly outline the costs associated with the retroconversion of *The American Archivist* into digital format following the model proposed in "Appendix VIII. *American Archivist* Retroconversion Report" of the Final Report of the SAA Task Force on Electronic Publications. This cost model is based upon how the Digital Research Library at the University of Pittsburgh would provide access to a full-text collection of this scale. It assumes substantial existing resources in the department, but that some additional resources would need to be purchased to mount, serve and host such a project.[8]

---

[8] This cost estimate does not include a service charge for hosting such a project; this document should not be considered a bid to do so.

Project Goals

The estimated costs are tied to the following project goals:
- Creation of digital page images and searchable text for past issues of the *American Archivist*;
- Creation of SGML-encoded text to capture structural characteristics of the *American Archivist*;
- Indexing of the encoded text to enable search and retrieval of the full-text and bibliographic information;
- Creation of a website to allow end-users to search, browse and print the page images of the *American Archivist*.

Project Costs

- Pre- and Post-Scanning Tasks
- Digitization of the *American Archivist*
- Technology and Infrastructure

Pre- and Post-Scanning Tasks

The work to achieve the project goals includes the following steps: data preparation, data conversion, data mounting, and data maintenance/preservation. It is estimated that it would take the DRL approximately 1,671 hours of labor to perform these tasks at a cost of $26,200.

STEP 1. Data Preparation	STAFF TIME
Acquire complete run of the *American Archivist* by SAA	--
Determine technical specifications and procedures	15 hrs.
Determine filing-naming scheme	2 hrs.
Determine intellectual access points, inc. defining document structural features, and anticipating how the *American Archivist* will be used online	30 hrs.
Determine any needs other than online presentation	2 hrs.
Design tracking database	10 hrs.
Total estimated time	**59 hrs.**

STEP 2. Conversion Process	STAFF TIME
Write scripts for automated processing of metadata spreadsheets and quality control	20 hrs.
Create metadata spreadsheets	500 hrs.
Proof metadata spreadsheets	150 hrs.
Pre-scanning data preparation	5 hrs.
Scan material by vendor (see Digitization)	30 wks.
Post-scanning data loading, image checking and rotation	16 hrs.
QC image content	330 hrs.
Pre-zone images for OCR	330 hrs.
OCR images	3 hrs.
Total estimated time	**1,354 hrs.**

STEP 3. Mounting Online	STAFF TIME
SGML creation and concatenation; move images and data to production server	3 hrs.
Collection configuration is entered into middleware system	1 hr.
Develop any custom search behavior	20 hrs.
Index SGML	3 hrs.
Test index for functionality	2 hrs.
Design custom Web pages by modifying templates	10 hrs.
Test Web pages for functionality	3 hrs.
Proof online content	200 hrs.
Public release of collection	1 hr.
Add collection metadata to local/national catalogs	5 hrs.
Make metadata available for harvesting via Open Archives Initiative (OAI)	10 hrs.
Total estimated time	**258 hrs.**

STEP 4. Maintenance and Preservation	STAFF TIME
Store original images on CD as delivered by vendor	--
Migrate files on the production server as necessary	--

Total Estimated Time: 1,671 hours

Digitization of the American Archivist

Based on the physical characteristics of the journal and the number of pages (see below), it would cost approximately **$6,150** for a scanning vendor to digitize the complete run of the *American Archivist* (note: this estimate is based on disbinding all issues for flatbed scanning). The digitization quote includes the creation of 1-bit TIFF 6.0 images of the text, 24-bit images of the 16 color plates (Vol. 63, Issue 1) and journal covers, and minimal post-scanning cleanup.

Scanning costs are based on the following characteristics of the *American Archivist*:[9]

Physical size of the journal	6.75" x 10"
Number of volumes	65
Number of issues	249
Total number of pages to be scanned	30,965
Number of pages that contain halftones	500*
Number of pages that contain color	16
Number of journal covers to be scanned in color	249*
Image cleanup to be performed	10 hrs.*
Shipping and handling	3 shipments

*estimates

---

[9] Includes all issues of *The American Archivist* published between 1938 and 2002, inc. Volume 65, Number 1, Spring/Summer 2002.

Technology and Infrastructure

Costs associated with mounting, maintaining, supporting and preserving the digital content are difficult to determine. Costs include: hardware and software, including the costs associated with yearly maintenance and licensing fees; disk space; and middleware (e.g., search engine; scripts that search, display, and browse the contents). It is estimated that a one-time cost of approximately **$35,000** would be necessary to purchase, configure and maintain a dedicated server running the middleware with a yearly maintenance cost approaching **$5,000**.

Summary of Project Costs

EXPENDITURES	COSTS
Pre- and post-scanning tasks	$26,200
Digitization of the *American Archivist*	$6,150
Technology and infrastructure	$35,000
Yearly maintenance	$5,000
Total	**$72,350**

APPENDIX IX
Content Submission Recommendation
Rob Spindler

SAA has traditionally received content for publication from authors via text in email messages, attachments to email, or documents and/or electronic media sent via s-mail. Content has typically been submitted in either Word® or WordPerfect® word processing files. All of these forms require handling by SAA staff and usually require several behind the scenes actions to retain the content, format it for publication, and complete publication agreements that address copyright.

The SAA TFEP has identified several types of content that fall within its mandate:
- *American Archivist* articles, letters to the editor, and Council minutes.
- *Archival Outlook* feature articles and news notes
- Annual Meeting program descriptions and presentations
- Education program course descriptions and curricular materials
- Monographic manuscripts
- Professional Opportunities and advertisements
- Bulletin Board
- Section newsletter content?
- Newly proposed forms of content (i.e. technical bulletins)

Automating the submission process presents several opportunities to streamline the process and minimize the potential for human error in document management. It also presents the opportunity to reduce the amount of return mail required for editorial work and rights management, and to manage the different actions required for managing different types of content. It also enables efficient aggregation and reaggregation of content into new and existing publication products.

An automated submission process can address several issues:
- Submission method
- Submission format
- Capture of descriptive and administrative metadata
- Rights management
- Editorial review and version control
- Content reformatting and archiving

We recommend that the SAA content submission process be directed toward the first four issues to begin, while automated editorial review, content reformatting and archiving functions could be added at a later date.

Submission Method:

SAA should develop a web-based submission site. This site would enable authors to visit a single page on the SAA website to deliver content and related metadata for all content submissions.

Submission Format:

Microsoft Word 98® and above probably offers the greatest flexibility in terms of subsequent conversion to production, archiving and delivery formats since it appears to support the cleanest automated conversions to XML amongst the most widely used word processing formats. We recommend it be identified as the preferred file format for submissions, while WordPerfect might be considered an acceptable format for text. Microsoft PowerPoint® may be established as the standard submission format for Annual Meeting presentations. These submission formats should be periodically reviewed and tested against new releases of the software (particularly the XP Office Suite) or other new software packages to ensure the cleanest migrations and conversions.

Word also offers the potential for embedding digital illustrations, but this presents some significant challenges in terms of file size and Internet file transmission since resolution has to be sufficient to support hardcopy publication quality content in many cases. SAA should probably identify jpeg format [cont. discussion] at a specified set of minimum pixel dimensions and resolution as the file format standard for imaged content. Until high speed transmission and email system limitations are increased illustrations may have to be submitted separately via hard media and s-mail, however authors will need direction on how they can identify place holders in their Word text for the images separately submitted. Submission standards for text and images should also include a protocol for file naming, which will be needed for automated control of the content at later stages.

Capture of Metadata:

The web-based submission format should enable authors to enter several pieces of administrative and descriptive metadata that would be used to populate a content control database and/or Dublin Core file header information. The DC file header information could eventually be used to support access via an OAIS server. A sample screen format for the submission interface could look like this:

SAA the society of american archivists

Authors Submission Site

Submission Help:

  FAQ

  Email to SAA Staff

SAA Copyright FAQ

File Format Standard

File Naming Protocol

Editorial Policies

SAA Privacy Policy

Author Name:

Last: [   ] **First:** [   ] **MI:** [ ]

Address:

Street: [   ]

City: [   ] **State:** [ ] **Zip:** [   ]

Country: [   ] **Phone:** ( ) **Email:** [   ]

Content Type: [   ]

Copyright Agreement:

I agree to generic agreement: Yes [   ] **No, please call me:** [   ]

Please mail monograph/journal article/curriculum agreement: [   ]

Attach File Submission: [   ]

Thanks for supporting the Society of American Archivists!

Data Dictionary for Author's Submission Site:

Author Name: All fields manually entered, no controls

Address: (needs to accommodate international submissions)
        Street: Manually entered, no controls
        City: Manually entered, no controls
        State: Drop down standardized abbreviations
        Zip: Drop down listings city/state specific
        Country: Drop down listings (standard?)
        Phone: Manually entered, area code parenthesis provided?
        Email: Manually entered

Content Type: Drop down menu of intended content type, values:
        Advertisements – generates pop up requirements box and invoicing
        Annual Meeting program descriptions
        Annual Meeting presentations
        American Archivist articles
        American Archivist letters to the editor
        American Archivist book reviews
        American Archivist SAA Council minutes
        Archival Outlook feature articles
        Archival Outlook news notes
        Bulletin Board
        Education program course description
        Education program curricular materials
        Monographic manuscript
        Professional Opportunities – generates pop up requirements box and invoicing
        Section newsletter – list valid SAA Sections
        Technical bulletin

Copyright Agreement:

"Generic agreement" links to full text of shrink wrap agreement applicable to most forms of news notes and other ephemeral content where copyright is not likely to be disputed by authors. "Yes" Box results are saved as metadata entry verifying agreement. "No" box provides option for discussion with SAA staff, and could result in pop up email box for questions about generic agreement.

"Monograph/Journal Article/Curriculum Agreement": Check here invalidates any entry in the "generic agreement" area and results in automated e-mailing of cover letter and standard copyright agreement. This form of agreement can be negotiated and requires handwritten signature submitted via return s-mail.

Attach File Submission: A Click in this box would result in pop up directory of the author's hard drive for attachment identification and uploading. This could be authenticated for correct file naming and extensions (i.e. Word document filename without the author last name and .doc extension is rejected).

Help Frame:

Submission Help:
  FAQ: Standardized textual instructions for completing the form. Could eventually be presented as online tutorial
  Email to SAA Staff: Email link to publications director or designate

SAA Copyright FAQ: Standardized textual explanation of SAA's generic and specialized copyright agreements drafted by SAA staff and reviewed by Council and SAA attorney.

File Format Standard: List of acceptable file format standards.

File Naming Protocol: Description of file naming protocol used to verify submission filename.

Editorial Policies: Standardized textual policies listed by content type.

SAA Privacy Policy: Standardized textual policy drafted by SAA staff and reviewed by Council and SAA attorney.

_____

Rights Management:

SAA should develop copyright agreements for its content that enable aggregation into the original intended publication and also reaggregation into other current and new products. A "shrink wrap" style agreement can be completed via a simple check box retained with the author metadata in the web form. This should be sufficient for most routine news and newsletter feature articles. TFEP recommends a more substantial hardcopy agreement for formal journal articles, monographs, and annual meeting presentations. This agreement may be negotiated with authors to allow them to retain rights and license content to SAA when necessary, but it should also enable the repurposing of this content for SAA's current and future needs.

Editorial Review and Version Control:

Initially TFEP recommends continuance of the current manual editorial review process whereby files are selected and e-mailed to editors and blind reviewers by SAA staff. When files are submitted the SAA server should automatically place those files in specially named directories tied to the "Content Type" field.[10] The author name presented in the filename (per file naming standard) and automatically generated date of submission should be sufficient to locate the submitted content in the server.

In future it should be possible to automate editorial work by establishing a secure website that editors and reviewers may enter and electronically mark documents with commentary. This functionality requires actively managed authentication to allow access for selected reviewers, and sophisticated version control that will require modification of file naming standards and specialized metadata collection. Word currently contains some version control features in its "Track Changes" functionality, but this may not be sufficient for our needs.

Content Reformatting and Archiving:

For the present TFEP recommends that content reformatting for production and archiving be handled manually, such that SAA staff select files and manually convert them to the required production, archiving and delivery formats, and store them in appropriately named subdirectories on the server. In future it should be possible to automatically conduct these reformatting activities in batches based upon "content type" and date of submission or a separate metadata element for date of publication maintained by the SAA staff.

rps: 5/4/2002

---

[5] The concept of "Content Type" has appeared in several different contexts in this report. It has been used to refer to file formats, genre, and intended dissemination vehicle (i.e. *American Archivist,* Annual Meeting Presentation). In the content submission context it was intended to address the need to automatically file content in subfolders associated with the intended dissemination vehicle. Retrieval by file formats should be enabled, particularly in connection with version control, and the issue of genre may have to be resolved in relation to intended dissemination vehicle. Continued discussion.

APPENDIX X
Long-Term Management and Preservation of SAA Electronic Publications
Cal Lee, August 2, 2002

1. General Considerations

This is a preliminary list of considerations for the preservation of SAA's electronic publications. The specific nature of these considerations will be depend heavily on details of the further recommendations from Council and implementation decisions made by SAA staff.

1.1 Capture and Storage of Distinct Digital Objects

SAA is moving toward a content management model for its web site. This implies treating components of documents, rather than entire documents, as the basic units of analysis. A given piece of content can become part of multiple publications, at varying levels of aggregation. As archivists, it will be important for us to identify those cases in which a cluster of components constitutes a publication that we want to "freeze" in time by taking a snapshot of its content and storing it in a repository outside of the live system. This should be addressed early on, since it will often not be possible to capture fixed views of dynamic content elements after the fact.

1.2 Repeatable Characteristics of Digital Objects

Given the continuous changes in the underlying technology, digital objects will be moved from one platform to another over time. The goal should not be low-level fixity, but instead repeatability of important characteristics over time. If a user sees a given SAA publication today, she should be confident that viewing it 10 years from now will yield a publication with similar characteristics, even if the hardware and software involved in its management and preservation have changed.

1.3 File Formats

SAA should establish conventions for formats to be used when setting aside content for long-term preservation. Ideally these formats will be based on open industry standards, rather than being dependent on one or a small number of proprietary systems. To minimize complexity, the number of formats should be kept relatively small. In some cases, however, SAA may wish to select more than one preservation format for a given set of publications, in order to hedge its bets on future support and availability of these formats.

1.4 Preservation Metadata

All electronic publications should include a minimum set of preservation metadata elements. These metadata are distinct from elements used for description, resource discovery or the administration of rights and services.

1.5 Linking and Naming

Each resource should be assigned an identifier that is consistent and globally unique. This will minimize the expense and complexity of long-term management of the resources, and it will also help to ensure the integrity of links from external sources (e.g. commercial search services, online public access catalog records, third party aggregators, citations in articles). Although the Task Force does not recommend that SAA immediately take on the issue of reference linking, it is important that SAA establish naming conventions that do not preclude such efforts in the future. At a bare minimum, directory and file naming should be consistent and should not make use of white space or other characters that cause problems in current operating systems.

A variety of strategies have been proposed for addressing the persistence of links,[11] but most include some sort of string to serve as an identifier, a specification of a domain within which the identifier is meant to apply, and some mechanism for resolving the identifier as a reference to a specific location within a storage system. Digital object identifiers (DOIs) have received considerable attention from publishers and librarians in recent years. There are also approaches that accept the URL as a fairly brittle linking mechanism, but take measures to accommodate for its limitations. The concept of "robust hyperlinks," for example, augments standard URLs with a word content-based lexical signature, i.e. a combination of text from the resource's own content that is assumed to be unique to that resource.[12] There are also a number of more simple practices that can minimize the software and format dependencies of links.[13] In general, the use of linking approaches that depend on specific software or proprietary formats should be avoided. Using links from PDF files, for example, should be used only in very limited cases, e.g. a PDF surrogate of a document, the original version of which included links to external Web resources.

1.6 Distinction Between Preservation and Dissemination Versions

It is important to distinguish between digital objects being set aside for preservation and those being disseminated directly to users. A given document, for example, might be stored in XML but served to users in PDF or HTML/XHTML.

1.7 Backup and Redundancy

No approach that SAA adopts should depend on the persistence of a single copy of a given electronic publication. Redundant storage is essential.

1.8 Third Parties and Licensing

It is quite possible that SAA will rely on third parties, not only for processing and dissemination of publications but also for ongoing preservation and access. Contracts and licenses should include clear provisions for preservation. They should also not transfer exclusive intellectual property rights to another party. Finally, there should always be a clear "exit strategy" in place, given that a vendor goes out of business, gets purchased by another company or changes its business model in such a way that no longer provides support for ongoing access to the publications in question.

2. Substantive Recommendations

2.1 Adoption of the Reference Model for an Open Archival Information System (OAIS)

The Reference Model for an Open Archival Information System has recently emerged as a common source of vocabulary and conceptual guidance for digital preservation efforts.[14] For the purposes of this Task Force, it is important to distinguish between the following three terms defined within the OAIS Reference Model:

- **Archival Information Package (AIP)**: An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.

---

[11] See http://www-personal.si.umich.edu/~calz/ermlinks/stan_rid.htm

[12] http://www.cs.berkeley.edu/~phelps/Robust/index2.html

[13] Kelly, Brian. "Guidelines for URI Naming Policies." *Ariadne* 31 (2002). http://www.ariadne.ac.uk/issue31/web-focus/

[14] "Reference Model for an Open Archival Information System (OAIS)," Washington, DC: Consultative Committee for Space Data Systems, Blue Book. Issue 1, January 2002, http://www.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf

- **Dissemination Information Package (DIP)**: The Information Package, derived from one or more AIPs, received by the Consumer in response to a request to the OAIS.

- **Submission Information Package (SIP)**: An Information Package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs.

According to the Reference Model, "ingest" is the process of accepting SIPs, preparing AIPs for storage, and ensuring that AIPs and their supporting descriptive information are established within the OAIS. It is important to note that an OAIS is charged specifically with preserving information and making it available over time. Most of SAA's processes and procedures for electronic publishing will fall outside of the scope of this model. When an author submits an article to SAA, for example, this transfer of an electronic file is distinct from the process of ingest into an OAIS. Ingest ensures that digital objects take the appropriate form for long-term preservation, including conformance to file format requirements, persistent naming and association of metadata. Ingest also focuses on the final version of a digital object as it was structured and formatted when published, whereas submission by an author to SAA will often involve preliminary drafts which will not be preserved over time.

2.2 File Formats

For each electronic publication, SAA should identify the format and elements required for the AIP, DIP and SIP. As mentioned earlier, it is possible that the best approach for some publications could be adoption of multiple AIP formats, so that the range of preservation and long-term access options will be maximized. The following is a set of tentative recommendations for each of the major content areas. Each is subject to further review and revision.

In order to facilitate long-term preservation and access, we recommend that SAA actively investigate the use of XML as the AIP format for electronic publications. XML could provide some serious preservation and repurposing advantages over formats such as Microsoft Word or even PDF, if used in a controlled enough way. One promising factor, which Brian Doyle pointed out to the Task Force, is that the latest version of QuarkXPress, the page layout program used by SAA for print production, supports the exporting of XML through a supporting application called Avenue.quark. It is important to investigate how seamless this process of generating XML from this product may or may not be. One of the primary advantages of PDF is that many applications can generate it natively to create a document that is visually similar to the version being used for print publication. The Council on Library and Information Resources (CLIR) informed us that they still rely on PDF, because cleaning up the HTML is very labor intensive. While PDF raises potential concerns about accessibility and the proprietary nature of its specification,[15] many have advocated the use of PDF as a preservation format.[16]

If the XML from Avenue.quark could be used to serve XHTML (an implementation of HTML that conforms to the markup conventions of XML) to browsers, that would be a terrific option, since it would allow for closer alignment of preservation and access formats. If the process requires a great deal of cleaning up and reformatting, however, it might not be a very scalable approach. In order to rely on XML, it is advisable to make use of formally defined document type definitions (DTDs) or schemas. Recent work on DTDs for the preservation of electronic publishing seems quite promising.[17]

---

[15] The specification is detailed and publicly available, but Adobe still owns the rights to it, which raises questions about their degree of future control over implementations based on the specification.

[16] See John T. Phillips, "Should PDF Be Used for Archiving Electronic Records?" *Information Management Journal* 35, no. 1 (2001): 60-63; John Mark Ockerbloom, "Archiving and Preserving PDF Files." *RLG DigiNews* 5, no. 1 (2001); Victorian Electronic Records Strategy (VERS), http://www.prov.vic.gov.au/vers/.

[17] See e.g. Inera Incorporated. "E-Journal Archive DTD Feasability Study." Harvard University Library, Office for Information Systems, E-Journal Archiving Project, 2002. http://www.diglib.org/preserve/hadtdfs.pdf

	Submission Information Package		Archival Information Package		Dissemination Information Package	
	Interim	Desired	Interim	Desired	Interim	Desired
Archival Outlook	PDF	XML	PDF	XML	PDF	XHTML
American Archivist (Born-Digital)	QuarkXPress	XML	PDF			XHTML
American Archivist (Retro-Converted)	TIFF and XML	TIFF and XML	TIFF and XML	TIFF and XML	GIF and HTML	GIF and XHTML
Monographs	QuarkXPress	XML	PDF	XML	PDF	XHTML or PDF
Annual Meeting Program	HTML	XML	HTML	XML	HTML	XHTML
Section Newsletters	HTML	XHTML	HTML	XHTML	HTML	XHTML
Annual Meeting Session Papers	Word PowerPoint	XML	HTML? Word/PowerPoint?	XML	HTML	

2.3 Preservation Metadata

The past couple years have seen a great deal of activity in the area of digital preservation metadata standards development. We recommend that SAA develop its own local metadata profiles that are based on this recent work. Specifically, we would recommend selecting a subset of the elements described in the most recent guidance from the OCLC/RLG Working Group on Preservation Metadata.[18]

2.4 Physical Storage Media

For those digital objects that SAA preserves in-house, we recommend the use of magnetic hard disks as the storage medium. Hard drive capacity continues to increase and become less expensive. They also have very fast read/write times and they frequently refresh the data, so "bit rot" is less of an issue than with some other media. If SAA chooses to retain CDs on their original media, it is important there be a plan in place for their transfer to another storage medium in a few years. They should also be stored in controlled conditions with appropriate labels and containers, rather than simply leaving them in sleeves such as those included in the backs of books. Other optical media such as DVD may eventually be viable media for the ongoing storage of electronic publications, but SAA should not yet rely on them for this purpose.

As described in section 1.7, storage of AIPs should always include redundancy. This will ideally involve mirror copies of all files in a physically location that is separated by several miles from the primary storage location. Off-site storage could take the form of additional magnetic hard drives or some other medium such as magnetic tape.

---

[18] "Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects," Dublin, Ohio: OCLC Online Computer Library, 2002, http://www.oclc.org/research/pmwg/pm_framework.pdf
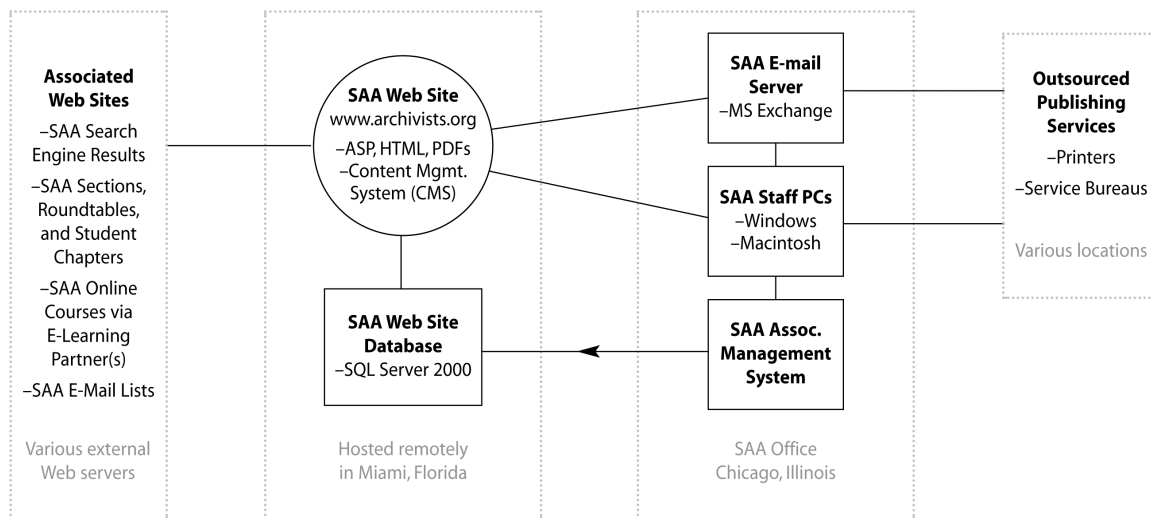
APPENDIX XI
E-Publishing and Technological Inrastructure
Prepared by Brian P. Doyle, SAA Webmaster/Graphic Designer
August 2002

This report examines recommendations of the SAA Task Force on Electronic Publishing (TFEP) subcommittees and provides analysis vis-à-vis SAA's current and prospective technological infrastructure. This infrastructure is comprised of multiple systems, servers, and databases, a fact that presents both opportunities and challenges for information networking and electronic distribution of SAA content. We begin with an overview of SAA's current systems.

SAA's Information Systems (Current Infrastructure)

Associated Web Sites
–SAA Search Engine Results
–SAA Sections, Roundtables, and Student Chapters
–SAA Online Courses via E-Learning Partner(s)
–SAA E-Mail Lists

Various external Web servers

SAA Web Site
www.archivists.org
–ASP, HTML, PDFs
–Content Mgmt. System (CMS)

SAA Web Site Database
–SQL Server 2000

Hosted remotely in Miami, Florida

SAA E-mail Server
–MS Exchange

SAA Staff PCs
–Windows
–Macintosh

SAA Assoc. Management System

SAA Office Chicago, Illinois

Outsourced Publishing Services
–Printers
–Service Bureaus

Various locations

As the diagram above illustrates, SAA information is distributed through various systems both within and outside of its headquarters in Chicago. As SAA's "Web Strategy Plan," drafted by Teresa Brinati and Brian Doyle in January 2001, explains[19], the integration of SAA's association management system (AMS)--i.e., the core in-house database that stores membership, events, publications, and accounting information--is a challenging prospect. However, it is precisely this integration that forms the basis for more sophisticated electronic publishing endeavors.

In the past year, SAA has made progress in developing a more dynamic, database-driven Web site. This has been accomplished through the establishment of an external Web database and the use of Active Server Pages (ASP), a server-side scripting technology. SAA is also in the process of creating a content management system (CMS). The CMS provides user-friendly, Web-based control panels that allow content administrators to author and edit Web content. Records are stored in the Web database and displayed online via ASP templates. Finally, SAA is putting in place a "data mirror" that will copy key information from AMS to the Web database. This one-way information stream will be refreshed on a 24-hour basis, providing access to critical data for important Web applications such as SAA member authentication, an online membership directory, a shopping cart system, etc. However, the current model does not accommodate automatic, Web-based modification to records in AMS. Updates will continue to be submitted via Web forms and e-mail and entered manually by SAA staff.

---

[19] (see www.archivists.org/governance/saa2001webplan.asp)

SAA currently produces in-house print publications on a Macintosh G3 computer. QuarkXPress is used for layout and page composition. Adobe Photoshop and Adobe Illustrator are used for creation and editing of images and graphics. Adobe Acrobat is used to convert files to PDF format. Mac-format Quark files continue to be the industry standard for print publishing and are universally supported by printers and service bureaus.

Electronic Submission of Content

SAA content is currently submitted by e-mail, e-mail attachments, and Web-based e-mail forms (e.g., the Job Ads Submission Form, membership forms, etc.) The *Online Directory of Archival Education* also uses a content management system, by which users directly modify their records in the Web database.

It has been proposed that SAA adopt a single online interface for submission of all SAA content (emphasis mine). However, the increasingly dynamic nature of the SAA Web site means that, in certain cases, content will not be submitted as a file, per se, but as raw data that will be inserted directly into the appropriate tables in the Web database. In the latter case, such data may "go live" immediately upon submission, or it may be flagged as pending for review and activation by a content administrator.

Therefore, a single all-encompassing submission form may be impractical to design. Rather, an integrated set of forms could be created with a single start page that will direct the user to the appropriate submission form for the task at hand. This would accommodate various types of submission, including direct-to-database as well as file attachment via a "browse your local disk" utility.

Content Delivery Modes

SAA content is currently disseminated via print, HTML, PDF, and Active Server Pages that pull information from the Web database. TFEP recommendations call for increased use of Web-based delivery via HTML and PDF, as well self-selected "content aggregations." Task force chair Rob Spindler argues that "the idea of aggregating and disaggregating content" is probably the most powerful concept that emerged from the TFEP discussions.

PDF is frequently proposed as a format for electronic publishing. In addition to it's widespread use, the strength of the PDF format is that is maintains the formatting of the original publication, which corresponds to Cal Lee's "fixity of content" idea (see App. IX). Conversely, it may not readily facilitate the chunking and repurposing (disaggregation) of content. Attention must also be given to the size of graphics within a document so that the resulting PDF is not too large for distribution via the Internet. Finally, creation of a PDF requires special software. Thus, dependence on this format would necessitate that all proposed content authors (e.g., staff, sections, roundtables, etc.) have access to this software.

HTML is another frequently proposed format. The strength of HTML is its universality. However, with multiple authors creating SAA content, formatting discrepancies can arise. The use of standardized templates and formatting styles, supported by ASP and cascading-style sheets, would help to alleviate problems (e.g., navigation difficulties, diffusion of the SAA brand, etc.) that can arise from a lack of uniformity. Thus, ASP should be encouraged as the new model over HTML. Guidelines and FAQs can assist content authors in creating such documents.

XML: An Emerging Model?

In sharp contrast to the simplicity of HTML and the fixity of PDF stands XML and a host of other acronyms that are often heralded as the future of the Web and publishing in general. It is worthwhile to note that several technology solutions currently employed by SAA are increasingly XML-compliant.

QuarkXPress, the page layout program used for desktop publishing, now offers an auxiliary program called Quark.Avenue (see www.quark.com/products/avenue/) that "lets [a user] extract content

from QuarkXPress documents and store that content as XML (and) import existing XML content directly into formatted QuarkXPress documents using the XML Import XTensions." SQL Server 2000, the software used for SAA's Web database, also supports XML.

The implication is that, if SAA remains on it's currently technological track, XML may emerge as a model for the seamless integration of print-based and Internet-based publishing. The details of such a model are beyond the scope of this report but further examination may be warranted.

Conclusion

The various reports of the task force contain many provocative and ambitious recommendations, some of which are already being offered or are "in the pipeline." For instance, the schedule of upcoming SAA workshops is currently maintained via the content management system, which automatically updates the respective registration form. The annual meeting program is also maintained via the CMS. The establishment of a data mirror between AMS and the Web database will also enable SAA to deliver on key recommendations such as member authentication for restricted-access content and development of an online membership directory.

Further analysis will be required to make implementation of certain sophisticated proposals a possibility. For instance, the aggregation and disaggregation of content will require a method for capturing and storing metadata. The content management system under development includes a "taxonomy" feature that may prove useful in this regard, however it is probably not as encompassing a solution as XML.

Most importantly, proposals for future electronic publishing endeavors must take into account SAA's current and anticipated technological infrastructure so that those solutions will cleanly interface with systems already in place.

APPENDIX XII
Draft Member Survey Questions Regarding E-Publishing
Scott Schwartz and the SAA Membership Committee

A. End User
- How long has the individual been a member of SAA?
- Is this an institutional or individual membership?
- If this is an individual membership, what is the dues category for this membership?
- How long has the individual worked in the archival profession?
- What is the individual's highest academic degree earned?
- In what type of archival repository does the individual work?
- How many professional staff work in this repository?
- Does the individual have special access needs? (addresses ADA issues)
- Does the individual us digital publications/resources as part of their daily archival routine?
- What types of digital publications are used daily by this individual?
- What types of traditionally printed publications are used daily by this individual?

B. Access
- Does the individual have access to a computer?
- Does the individual have access to the Internet?
- Will access to digital publications be exclusively through the SAA website or will other digital formats be made available?
- Will SAA provide secured access to its digital publications through its website and how will this secured access be maintained?
- How would digital publications affect traditional ILL processes?
- Will traditionally printed publications remain available to those without digital access?

C. Content and Format

- Will Archival Outlook and Employment Bulletin be available only in digital format?
- Should there be retroactive conversion of all or select issues of Archival Outlook?
- Will Section and Roundtable newsletters be available only in digital format?
- Will Section and Roundtable newsletters be indexed?
- Who will responsible for producing and maintaining these newsletters?
- Will digitizing a complete run of the American Archivist benefit the archival profession?
- Would a printable digital copy of American Archivist or a specific article from the Journal be beneficial to the individual?
- Will an indexing of a complete run of the American Archivist benefit individual and institutional members?
- Should archival monographs be made available in digital format?
- Should student poster sessions, papers, and other forms of preliminary research be made available to membership in a digital format?

D. Membership Benefit and Cost

- Would providing on-line meeting and training registration be considered a new membership benefit?
- Would providing on-line access to a complete run of the American Archivist be considered a new membership benefit?
- Would other types of digital publications be considered a new membership benefit? (Need to specify)
- Would the added financial costs of digital production be offset by current revenue generated from membership dues?
- Would an individual be willing accept an increase in membership dues to defray added financial

costs possibly associated with the creation of digital versus paper productions?
- Who will be responsible for doing the added work of creating digital publications produced by the SAA Office?
- Who will be responsible for doing the work of creating digital publications (e.g., section newsletters) not produced by the SAA Office?
- How will resources that are currently devoted to traditional types of publication be diverted to produce digital publications and what type of impact will this have on types of publications that will not be digitized?
- What are the cost savings, in terms of production and mailing, by moving from a traditional style of publication to a digital format?