Reframing Collections for a Digital Age: A Preparatory Study for Collecting and Preserving Web-based Art Research Materials

Final Report
April 12, 2013

Objectives

The objective of the study was to identify the organizational, economic, and technological challenges posed by the rapidly increasing number of web-based or "born-digital" resources that document art history and the art market. Building on previous initiatives, the results of the study are intended to allow the New York Art Resources Consortium (NYARC) and other libraries specializing in art research to determine the best way to acquire, preserve, and provide unified access to these unique and often ephemeral materials. A further objective was to evaluate NYARC's existing library systems, workflows, and policies to address requirements of next-generation systems that will enable NYARC to forge new models in building collections in a digital age.

Deliverables

The study was structured as a series of three separate outside consultancies, each one informing the subsequent one. Specific goals focused on a set of deliverables as follows:

- Environmental and literature surveys of web-based and born-digital publications to determine where the tipping point from print to digital may lie for specialized art resources
- Input from the user community on current use and needs for web-based resources in art history
- Recommendations for materials to collect
- Recommendations for best methods of web archiving
- Identification of technical and other collaborative partners
- Review of existing technical infrastructure and recommendations for technical changes to support digital capture, unified resource discovery, access, and preservation
- Sustainable solutions for rebalancing budgets and staff during a hybrid period of managing analog and digital resources
- Guidelines for intellectual property management procedures and applications of fair use that will determine levels of access
- Preparation of systems needs analysis and requirements for RFP
- Dissemination of results to the art libraries community at large

Accomplishments

Three consultants with expertise in libraries and technology were engaged. Over the course of the study, each consultant met with the NYARC directors and with one another to share findings so that each report could build on the previous ones. Legal advice was also sought, resulting in a set of guidelines for expanding the web archiving projects begun at the Frick. Basecamp project management software was set up and contains records of meetings, contacts, discussions, citations, blogs, joint grant applications, evaluation forms, files, and other documentation that will be relevant as we move towards a defined program for web archiving.

Our first consultant, Doralynn Pines, former administrator at The Metropolitan Museum of Art, conducted a field assessment of the state of publishing in the arts to determine where the "tipping point," or the moment of critical mass, for the transition of analog to digital formats lies for specialized art resources. She interviewed directors at market leaders including Pace Galleries, Sotheby's, the Oxford, Cambridge, and Yale University presses, as well as leading museum publishers, and concluded that in the world of traditional publishing the tipping point remains on the horizon because of concerns about rights issues, while organizations that collect contemporary art may be at the cusp of going chiefly digital in their publishing efforts. She also conducted focus groups with curators, conservators, educators, archivists, and new media staff at the three NYARC parent institutions to determine researcher requirements for using and citing digital works. To no surprise, she confirmed widespread confusion among curators and scholars regarding e-formats and permanence of digital resources. Priorities did emerge, however, including the need for permalinks to citations, and a need to capture iterations of web-based exhibitions, catalogues raisonnés, and auction catalogs. Invariably there were questions about responsibilities and costs, and where the boundaries of collecting lie for works of art that are born-digital. Gail Truman, the second consultant, used the focus group results as the basis of her report, which used case studies to provide scenarios of what to collect, the best methods for web archiving, and what partners (technical, publishing, other research libraries or preservation consortia) to work with. The third consultant, Lily Pregill, used conclusions and recommendations from the previous reports to assess the technology and workflows required to support a web archiving program at the NYARC libraries.

The following reports and documents were delivered to the group during the grant period:

- *Reframing Collections for a Digital Age 1: The Tipping* Point / Doralynn Pines. July 4, 2012; Addendum August 14, 2012 (22 p.)

- *Reframing Collections for a Digital Age: Focus* Groups / Doralynn Pines. July 23, 2012, (28 p.)

- *NYARC, Reframing Collections for a Digital Age: Report from Consultant No. 2* / Truman Technologies, September 11, 2012, (32. p.)

- IVAAG Web Archiving Survey Responses / Lily Pregill, December 10, 2012.

- *Guidelines for Collection and Preservation of Online Materials (Draft)* / Robert Clarida, January 16, 2013, (5 p.)

- *Reframing Collections for a Digital Age: Infrastructure & Workflow Report* / Lily Pregill, January 25, 2013, (42 p.)

- *Reframing Collections for a Digital Age: A Preparatory Study for Collecting and Preserving Web-based Art Research Materials: Summary of Consultants' Recommendations* / Deborah Kempe, January 2013 (5 p.)

In support of the goal to disseminate the consultants' findings to the art library community, Deborah Kempe, the Principal Investigator, delivered two presentations during the grant period:

> *Bringing Light to the Black Hole: Experiments in Archiving Born-Digital Materials*, Art Libraries Society of North America 40th Annual Conference, Toronto, Ontario, March 31, 2012.

Reframing Collections for a Digital Age: The Challenge of Web-Based Art Resources,
Artlibraries.net Meeting, Paris, France, September 27, 2012 (and the NYARC Annual Meeting, February 19, 2013)

Discussions of legal issues with regard to web archiving and access were initiated and are ongoing with Robert Clarida and Kenneth Crews, two leading lawyers in the field of intellectual property and copyright. The areas of discussion include criteria for selection of sites to harvest, terms of use, schedules for determining status of urls, legally relevant technology issues, and means/extent of access.

NYARC has moved to adopt many of the recommendations made by the consultants and legal advisors, including:
- Selection of Archive-It as web archiving tool of choice
- Consolidation of Archive-It subscriptions at the Frick
- Adjustment of existing web capture based on emerging legal guidelines
- Using graduate student interns for quality assurance of web capture
- Preparation of lists of targets for web archiving, prioritized by importance to the community or danger of impermanence
- Becoming a member of the National Digital Stewardship Alliance (NDSA)
- Establishing system requirements in preparation for RFP
- Ongoing communication with Hanzo Archives about non-profit models for web archiving
- Ongoing communication with library service platform developers and advocacy for centralized web archiving solutions in their products
- Determining DuraCloud requirements for a funding proposal to the Mellon Foundation
- Drafting a policy document with goals and objectives for the NYARC web archiving program
- Investment in training for staff to learn more about web archiving and managing multiple metadata formats
- Appointment of a NYARC representative to the Archive-It Metadata Working Group

Above all, the primary message in the reports of all three consultants was that the innovative and sustainable solutions to address the challenges posed by web-based art research materials will require deep collaboration among the auction houses, publishers, and others who produce them, the libraries that collect them, and the technical companies that make it possible.

The NYARC Directors and consultants also made contact with many other groups active or with an interest in web archiving:

- Deborah Kempe and consultant Gail Truman participated in the groundbreaking Web Archiving Policies and Practice in the US: 2012 Summit at Columbia University Libraries.
- Hanzo Archives provided their first demo to non-profit groups at a December 2012 meeting hosted by NYARC at the Museum of Modern Art.
- Frequent meetings and phone calls were held with web archiving experts at the Library of Congress, British Library, Bibliotheque national de France, and Columbia University Libraries.
- Lily Pregill attended the Archive-It Partner Meeting in Annapolis, MD, December 2102.
- NYARC became a partner with the Information Science Department of Old Dominion University, VA, a leader in web archiving research, to participate in two proposals to the NEH, for work with a self-archiving tool that could be used by artists, and for data visualization methods.
- Two paid interns from Pratt School of Information & Library Science worked on web archiving as part of a three-year project funded by the Institute of Museum and Library Services.

- NYARC became a Strategic Research Partner for the development of a copyright duration tool being developed at Tulane University.
- A web-archiving survey of the IVAAG (Ivy Art and Architecture Libraries Group) was conducted and presented at Princeton University in November, 2012.
- Stephen Bury, Andrew W. Mellon Chief Librarian of The Frick Collection, was appointed to the ITHAKA Research and Support Services Art Historians Advisory Board. The board's investigations include born-digital resources in art historical research.

Challenges

For all of the progress made over the course of our study, it often feels that the more we learn about the ever-evolving nature of web publishing, the larger the questions and obstacles loom. Indeed, the challenges remain daunting. No library entity, whether a single specialized research institute, a small consortium such as NYARC, a large state university facility, or a well-endowed Ivy League library can alone tame the onslaught of born-digital media, even when narrowed to the universe of art scholarship. Mobilization on the national level does not look promising in the near term, as the U.S. Copyright has for the foreseeable future exempted online-only works from the requirement of mandatory deposit.[1]

Staffing: The Preparatory Study anticipated that staffing requirements for collecting born-digital material would be met by rebalancing existing staff. Reports by all three consultants, however, were explicit that, due to the hybrid nature of the publishing landscape for the near future, and the complex legal decision-making that will be required to implement a fully functional web archiving program, additional staff will be required. Graduate students may offer a sustainable near-term solution. As information slowly shifts from print to digital, and expected improvements to software to automate manual review emerge, staff duties may be incrementally shifted to responsibilities for digital management. These duties may and should be shared by the NYARC partners to achieve cost effective results and consistent policies and procedures. Nevertheless, it is clear that additional staff hours over and above existing levels will be required to manage hybrid collections during this period of transition.

Scale: Given an unlimited number of potential producers of born-digital material, a workable method to distribute responsibility and costs for web archiving remains elusive. Most programs today operate on a boutique level and no effective cooperative models are in place. This is an area of opportunity for NYARC to lead the way for art research collections.

Infrastructure: Costs of web archiving tools and storage solutions remain high. Library services platforms have not yet evolved to support web archiving workflow or to fully align with institutional repositories, DAMS, or preservation platforms. It may be expected that prices will go down when web archiving is more widely adapted, but there is currently little room in the NYARC budget to support these new costs.

Description and Access: While web-archiving addresses data capture at a meta level, distinctive documents embedded in those sites are the equivalent to printed documents described in online catalogs. NYARC and other libraries struggle to determine the proper level of granularity for description of and access to born-digital resources. Archival methods of creating parent-child relationships are possible but may introduce a new level of costs. Automated tools may emerge to alleviate manual entry.

---

[1] http://www.copyright.gov/fedreg/2010/75fr3863.pdf

Legal issues: Our chief legal advisor recommends very restrictive policies. The irony of locking down information that in its original form had unrestricted and unlimited access is not lost on us, and we will continue to pursue less restrictive options, including working under a liability umbrella of a large university that has an affirmative stance on fair use. Access beyond library walls and avoiding dark archives must be an objective, or we have failed in our mission.

Project Personnel

There were no changes of project personnel on the grant.

Publications

Aside from the reports listed as accomplishments, there were no publications associated with the study.

Intellectual Property

This grant was not subject to an intellectual property agreement.

Future Plans

As a result of the intensive investigations that informed this study, the NYARC group is reasonably certain that we have the most accurate assessment of the current landscape of web-published art documentation that exists in the art information community. Our colleagues in this field are awaiting information on the final results and conclusions of the grant, and we plan to share these results in April at the ARLIS/NA annual conference and through the NYARC website. A process for partners to participate in a distributed model of building digital collections will be introduced at the ARLIS/NA conference.

Collaborations are crucial and will be incorporated in our strategic plan and roadmap for born-digital collections at NYARC. See Attachment 1 *"Collaborations: Ongoing and Potential."* The NYARC partners, with a record of successful cooperation in building new technological and service models, are uniquely positioned to forge new models in building collections in the digital age. We foresee an achievable set of deliverables for an implementation phase. Among them: integration of Archive-It with a commercial library services platform; the creation of community-specific guidelines for cataloging web-based documents about art; public/private partnerships to develop automated tools for discovering new sites and detecting trends in scholarship; and a cost-effective shared repository environment.

Given the frequent exchange of information on web archiving projects between NYARC and Columbia University Libraries, there is a mutual desire of the two entities for collaboration on next stages. At a recent meeting with James Neal, Columbia's Vice President for Information Services and University Librarian, and Robert Wolven, the Associate University Librarian for Bibliographic Services and Collection Development, details of the goals of Columbia's recently received Mellon grant for web archives were discussed.

Our study confirmed that while programs of web archiving are increasing at large research institutions, there is as of yet no organized effort to collect and provide access to free and commercial art historical resources published on the web. Archiving and managing these kinds of digital documents will be the focus of the 2013 NYARC's Strategic Planning retreat, where we will further explore ways of making web archiving scalable to the art research community, while at the same time taking advantage of an infrastructure that is growing more standardized as web archiving becomes more widely adopted by large

university libraries. Our strategy is to convey the message that the capture, access, and preservation of born-digital art research is essential for the future and that NYARC has an essential role in meeting this mission-critical objective. The only unacceptable option is to do nothing.

"As long as libraries and archives remain stymied by the plethora of issues involved in archiving born-digital content and the daunting prospect of having to 'do it all,' their progress will continue to be slow. Inaction may impede the course of research and contribute to the loss of important content. By leveraging specialized resources and expertise across the archival and library communities, individual archives and libraries are far more likely to be able to keep up with the onrush of born-digital content and actively further the course of research."
--Ricky Erway, OCLC Program Officer. Swatting the Long Tail of Digital Media: A Call for Collaboration—Sept. 2012 http://www.oclc.org/content/dam/research/publications/library/2012/2012-08.pdf