Reframing Collections for a Digital Age: The Challenge of Web-Based Art Resources

Artlibraries.net General Meeting
Paris, September 27, 2012

Deborah Kempe
Frick Art Reference Library
• International and national efforts
• Current state of the web
• NYARC as a case study
• Pilot Project to archive auction sites
• Reframing Collections Study Grant
• Recommendations
• Collaboration
WEB ARCHIVING POLICIES AND PRACTICE IN THE US: 2012 SUMMIT

A meeting of US institutions engaged in harvesting and archiving web content, to explore high-level programmatic issues of objectives, scope, policies and methods for this work.

Columbia University Libraries has invited participants from over 20 US academic and research libraries to share perspectives on topics relevant to their own experience, including:

- institutional objectives in developing a web archive
- assumptions about the long-term sustainability of local web archives
- the way in which web collections are scoped
- policies regarding permissions and control of content
- assumptions regarding the expected uses and users of archived content
- how those assumptions affect decisions about presentation and access

As an outcome, we hope the meeting may lead to an articulation of a small number of model programs for web archiving, and development of "best practices" for documenting program elements.
The Isamu Noguchi Catalogue Raisonné is committed to documenting Isamu Noguchi's complete oeuvre including sculptures, drawings, models, architectural spaces, stage sets, and manufactured designs. For information on the configuration of artwork entries see the Guide to Entries.

While research for The Isamu Noguchi Catalogue Raisonné is ongoing, the current list of artworks includes both published and research-pending entries. New additions and changes will be made on an annual basis until research is complete. Learn more.

Browse All Artworks »
**Roy Lichtenstein Chronology**

By Clare Bell (with additional content provided by the Lichtenstein Foundation)
copyright The Solomon R. Guggenheim Foundation 1993/1998

1923

Oct. 27. Roy Fox Lichtenstein is born in Manhattan at Flower Hospital on

http://www.lichtensteinfoundation.org
NOT

Why Archive the Web?

BUT

How to Archive the Web?

Who Archives the Web?

Who Pays for Archiving the Web?

How do People Navigate Web Archives?
"Right now, we're 100 percent ready to archive the way the Web was 10 years ago."

Are we confused yet??
Distinguished auction house transforms its print catalogues for the iPad using Adobe® Digital Publishing Suite, Enterprise Edition

Founded in 1744, Sotheby’s is one of the world’s oldest and most venerable auction houses. With salesrooms in New York, London, Hong Kong, and Paris, Sotheby’s sells a diverse array of rare and extraordinary objects including contemporary art, Chinese ceramics, magnificent jewels, historical artifacts, musical instruments, and wine.

Every year, the company conducts more than 300 auctions. Collectors are able to bid in person in the company’s salesrooms, over the phone, or online. Print catalogues published for each auction are deeply valued by Sotheby’s clients. Produced according to the highest standards of scholarship and design, they enable collectors to browse the lots in an auction and learn about them—their provenance, their art historical significance, their value. Sotheby’s catalogues are available for purchase and are distributed to clients worldwide approximately one month before an auction. During the week prior to an auction, Sotheby’s holds an exhibition, providing the opportunity for collectors to view items in person.
We are straddling two worlds: physical and digital
Uh-oh ...here today, gone tomorrow.

Links to nowhere

<p>| TITLE | 245 00 Welcome to PumaPaint!|h[electronic resource] :|ba Java telerobotic interface for painting across the Internet. |
|-------|-----------------------------|
| NOTE | 500 Title from title screen (viewed Dec. 10, 1998). |
| NOTE | 538 Mode of access: Internet via the World Wide Web. |
| NOTE | 520&quot;Project background: Well, it all started with Dr. Matt Stein's robot. Dr. Stein controls a Unimation Corp. Puma 760 robot in his lab at Wilkes University (2 hours north of Philadelphia). He needed a platform independent interface to control the robot, so that it could be used to demonstrate control topics to a wide audience. His chosen task: painting. So, Dr. Stein talked to Pete DePasquale whom he worked with at Sonalysts, Inc. for a short time. Pete started thinking about using the Java language to create a web-based platform to control the robot, so he went to Dr. John Lewis at Villanova Univ. (where Pete was a graduate student), and told him the idea. Dr. Lewis loved it, and we recruited Lara Blatchford, also a graduate student at Villanova. Together, they went to see Dr. Stein's robot at Wilkes at [sic] talk about the interface. While they were there, Pete took some photos of the robot in action. Then they returned to Villanova and started designing the interface&quot;--title screen. |
| SUBJECT | 650 0 Robots in art. |
| FORMAT | 655 0 Web sites. |
| ADD.AUTOR | 700 1 Stein, Matthew. |
| MARC INFO 856 4 | uhttp://pumapaint.rwu.edu/index.html |</p>
<table>
<thead>
<tr>
<th>Type</th>
<th>Code</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE</td>
<td>245 10</td>
<td>International Museum of Cartoon Art [electronic resource].</td>
</tr>
<tr>
<td>IMPRINT</td>
<td>260</td>
<td>Boca Raton, Fla. : International Museum of Cartoon Art, 1997-</td>
</tr>
<tr>
<td>NOTE</td>
<td>500</td>
<td>Title from title screen (viewed Apr. 28, 1999).</td>
</tr>
<tr>
<td>NOTE</td>
<td>538</td>
<td>Mode of access: Internet via the World Wide Web.</td>
</tr>
<tr>
<td>NOTE</td>
<td>520</td>
<td>&quot;To make accessible to the general public the largest collection of original historic and contemporary cartoon works in the world and to continually acquire such works.&quot; --Mission</td>
</tr>
<tr>
<td>SUBJECT</td>
<td>610 20</td>
<td>International Museum of Cartoon Art (Boca Raton, Fla.)</td>
</tr>
<tr>
<td>SUBJECT</td>
<td>650 0</td>
<td>Caricatures and cartoons</td>
</tr>
<tr>
<td>SUBJECT</td>
<td>650 0</td>
<td>Museums</td>
</tr>
<tr>
<td>FORMAT</td>
<td>655 0</td>
<td>Web sites.</td>
</tr>
<tr>
<td>MARC INFO 856 4</td>
<td></td>
<td><a href="http://www.cartoon.org/home.htm">http://www.cartoon.org/home.htm</a></td>
</tr>
</tbody>
</table>
What is Archive-It?

• “Archive-It,” a subscription service from the Internet Archive, allows institutions to harvest and preserve digital collections

• Display is through the IA Wayback Machine

• Utilizes open-source tools: Heritrix as webcrawler, NutchWAX for search, SOLR for indexing

• 190 partner institutions, in 17 countries, with 1,954 collections
Adventures in Web Archiving

Capturing born-digital content from auction house websites

2010 Pilot Project
Pilot Auction House Seeds

- Bonhams www.bonhams.com
- Dreweatts www.dnfa.com/
- International Auctioneers www.internationalauctioneers.com
- Tajan www.tajan.com
- Pandolfini www.pandolfini.it
- Günther Kunstauktionhaus www.dresden-kunstauktion.de/
- Nagel Auktionen www.auction.de/
- auction.fr www.auction.fr/
- R. W. Oliver’s www.rwolivers.com/
- Hosane www.hosane.com
- Heritage Auctions www.ha.com
Initial Questions

• What limitations does the Archive-It crawler have?

• How deeply are sites crawled?

• Can we crawl selectively within a site? At what level can we curate the collection?

• What is the best frequency for crawling auction house sites? How is that determined?

• Do we bump into any access issues regarding authentication for archived content?
What did we find?

• Diversity of formats: PDFs, XML, e-Catalogs, embedded audio and video

• Dynamism of continuing resources

• Some sites forego legacy formats in favor of a dynamic, comprehensive, database

• Some sites capture well, others not at all
Problems encountered

• Overwhelming data affecting crawl efficiency
• Robots, crawler traps, password protection
• Many formats do not capture well or at all
• Content can be scoped but not deleted
• Maintenance needed
• Pushing the IP limits, no built in permissions
What is on the horizon?

©Andreas Gursky, Rhein II
Reframing Collections for a Digital Age: A Preparatory Study for Collecting and Preserving Web-Based Art Research Materials

- $50,000 from the Andrew W. Mellon Foundation
- 12 months, began February 2012

New York Art Resources Consortium The Andrew W. Mellon Foundation
Expected Outcomes and Benefits

(1) Find the ‘tipping point’ from analog to digital

(2) Recommendations to NYARC as to: what it should collect; best methods of web archiving; what partners to work with; how to address intellectual property, ethical and access issues

(3) Recommendation of appropriate changes to technical infrastructure to support digital capture, resource discovery, preservation and access

(4) Preparation of grant proposal for technical infrastructure upgrading

(5) Dissemination of results to the art libraries community
RECOMMENDATIONS

• Use Archive-It as the web archiving tool
• Establish incremental growth of collections beginning with high priority or high risk materials
• Join the National Digital Stewardship Alliance (NDSA)
• Develop a tool for open nominations for site selection
• Investigate ways to further automate metadata creation
• Use students for Quality Assurance
• Develop levels of restricted access
• Develop and maintain a framework of notifications and requests
• Large doses of collaboration
COLLABORATION WILL BE REQUIRED FOR THE FUTURE OF ART BIBLIOGRAPHY

“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.”

http://www.oclc.org/content/dam/research/publications/library/2012/2012-08.pdf
Come along for the ride...
Links to resources cited, and other useful information on born digital content

http://spectrum.iee.org/telecom/internet/a-memory-of-webs-past/0
http://www.technologyreview.com/view/429274/history-as-recorded-on-twitter-is-vanishing-from/
http://netpreserve.org/ website of the International Internet Preservation Consortium (IIPC)
http://nyarc.org/content/reframing-collections-digital-age, blog posting by Stephen Bury, June 18, 2012
http://www.loc.gov/webarchiving/ Library of Congress Web Archiving
http://www.lichtensteinfoundation.org Lichtenstein Foundation