Bringing Light to the Black Hole: Creating Web-Based “Born Digital” Collections in Art Libraries

III Seminar on Museum Information Services, São Paulo, Brazil
November 27, 2014

Deborah Kempe
Chief of Collections Management & Access
Frick Art Reference Library, New York, NY USA
kempe@frick.org

Pebbles around a hole, Kinagashima-Cho, Japan (1987). Photo by Andy Goldsworthy
“Going forward, one of the biggest challenges scholars and curators of contemporary art and architecture face currently, and will increasingly face, is how to store, retrieve, and investigate born-digital materials.”

James Cuno, President and CEO of the J. Paul Getty Trust, How Art History is Failing at the Internet (November 19, 2012)
OVERVIEW

Background and evolution

2014/2015 NYARC born-digital program

Staffing and collaboration

Tools we’re using

Creating and describing collections

Building a sustainable program
FRICK ART REFERENCE LIBRARY, ESTABLISHED 1920

½ million books, 1.2 million photographs, and growing number of e-resources that relate mainly to fine arts from the fourth to the mid-twentieth centuries by European and American artists.

Known internationally for its rich holdings of auction and exhibition catalogs, the Library is a leading site for collecting and provenance research.

Archival materials and special collections augment the research collections with documents pertaining to the history of collecting art and of Henry Clay Frick's collecting in particular.
Center for the History of Collecting
Established at Frick Art Reference Library in 2007
What is NYARC?

The New York Art Resources Consortium (NYARC) consists of the research libraries of three leading art museums in New York City: The Brooklyn Museum, The Frick Collection, and The Museum of Modern Art. With funding from The Andrew W. Mellon Foundation, NYARC was formed in 2006 to facilitate collaboration that results in enhanced resources to research communities. Our goals include:

- Improve access to art research resources through technology
- Advance the scholarly, educational and cultural enrichment missions of the three museums
- Provide leadership in the development of innovative and model information services programs

Press & Publications

- Reframing Art Resources for the Digital Age: An Interview with Stephen Bury & Lily Pregill of the New York Art Resources Consortium
- The Signal: Digital Preservation
- NYARC Newsletter

Over 1 million records
Find that missing piece
Stars from our collections
Libraries working together
What do we mean when we say “Digital Black Hole?”
What we mean when we say Digital Black Hole
What we mean when we say Digital Black Hole

Content being produced

Content being archived
A web archive is a collection of archived URLs grouped by theme, event, subject area, or web address.

A web archive contains as much as possible from the original resources and documents the change over time. It is a priority to recreate the same experience a user would have had if they had visited the live site on the day it was archived.
Historical precedents

European leaders

American innovators
Thank you, Brewster Kahle!
COLLABORATION WILL BE REQUIRED FOR THE FUTURE OF ART RESEARCH

The Internet Archive is often thought to be archiving "the entire web" but in reality it is just a slice of what's available.

It is important for libraries and archives to also select and create collections of web content.

from Library of Congress Web Archiving FAQs
Song Dong // Regenerate
30/08 to 15/11/14

• 04.09.14 // David Medalla & Adam Nankervis // Inhotim // Minas Gerais
• 28.08 – 29.08.14 // Laerte Ramos // Casa das Onze Janelas // Bairro
• 25.08 – 25.10.14 // Laerte Ramos // Itaú Cultural // São Paulo
• 19.08 – 19.10.14 // Túlio Pinto // Felipe Moraes // IEA // Porto Alegre
• 08.08 – 11.10.14 // Túlio Pinto // SARP – Salão de Arte // Ribeirão Preto

+ mais noticias
NOT

Why Archive the Web?

RATHER

How to Archive the Web?

Who Should Archive the Web?

Who Pays for Archiving the Web?

How do People Navigate Web Archives?
ARCHIVING BORN DIGITAL CULTURAL HERITAGE OBJECTS-
NYARC’s EVOLUTION

2010-2013: Archive-It partnership for pilot studies and intern projects
Capturing born-digital content from auction house websites
2010 Pilot Project with Met’s Watson Library

Sean Leahy, Pratt MLIS Student Intern, Principal Investigator
PRESERVING WEB-BASED AUCTION CATALOGS

Gretchen Nadasky
M-LEAD-TWO Program
Frick Art Reference Library
Pratt Institute, LIS – 698
Dr. Tula Giannini
Spring 2013

Source: Dorotheum.com
A survey of web-based art resources with findings applicable to FARL electronic records collection development

Alison Rhonemus, LIS 698, Seminar and Practicum, Dr. Tula Giannini

Frick Art Reference Library
Deborah Kempe, Chief, Collections Management & Access
Web Survey and Collection Development
Archiving Auction House Websites

Ralph Baylor and Dana Hart

FARL Book Department Internship Project, Summer 2013
ARCHIVING BORN DIGITAL CULTURAL HERITAGE
OBJECTS-
NYARC’s EVOLUTION

2010-2013: Archive-It partnership for pilot studies and intern projects

Grant support from The Andrew W. Mellon Foundation:

2012-2013 planning grant ($50,000) ‘Reframing Collections for a Digital Age’
The Tipping Point from Print to Digital in Art History has been slow to approach
2013 PLANNING GRANT RECOMMENDATIONS

• Use Archive-It as the web archiving tool
• Plan incremental growth of collection
• Develop an open nominations tool
• Establish a permissions framework
• Join the National Digital Stewardship Alliance (NDSA)
• Look for ways to further automate metadata creation
• Enlist students into the program, especially for quality assurance
• Collaborate, Collaborate, Collaborate
ARCHIVING BORN DIGITAL CULTURAL HERITAGE OBJECTS -
NYARC’s EVOLUTION

2010-2013: Archive-It partnership for pilot studies and intern projects

Grant support from The Andrew W. Mellon Foundation:

2012-2013 planning grant ($50,000): ‘Reframing Collections for a Digital Age’

2013-2015 grant-funded program ($340,000): ‘Making the Black Hole Gray: Implementing the Web Archiving of Specialist Art Resources’
PROJECT OBJECTIVES

- During the 2-year grant period NYARC seeks to implement a program for capturing, making accessible, and preserving websites for art research

- Harvest and catalog approximately 2 TB of WARC (Web ARCHive file format) files from websites

- Implement a Discovery Layer

- Workflow documentation

- Develop and share best practices with the community
- NYARC institutional Websites
- Auction Catalogs
- Artists’ websites
- Catalogues Raisonnés
- NYC Galleries
- Provenance & Restitution
The Isamu Noguchi Catalogue Raisonné is committed to documenting Isamu Noguchi's complete œuvre 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 »
Web Archiving
Workflow
Elements

- Nomination/
  Selection
- Permissions/
  Tracking
- Harvesting & Quality Assurance
- Description/
  Access
- Long-term Preservation
Using technology in support of goals
ARCHIVE-IT  https://archive-it.org/
• Annual Subscription Service
• Crawls, harvests and hosts web content, using Open Source tools and standards developed and maintained by the Internet Archive:
  • Heritrix web crawler /+ Umbra
  • Nutchwax search engine
  • Wayback Machine browser
  • WARC files, ISO Standard
  • Latest version is 5.0
More than 330 partners in 16 countries use Archive-It

- College and University Libraries
- State Archives, Libraries, and Historical Societies
- Federal Institutions
- Non-governmental Organizations
- Public Libraries, Cities and Counties
- Museums and Art Libraries
Art museums and art libraries using Archive-It

- Hagley Museum and Library
- New York Art Resources Consortium (NYARC)
- Arab American National Museum
- New Museum
- Smithsonian Institution
- National Gallery of Canada Library and Archives
- National Museum of Women in the Arts
- Artist Files Special Interest Group (ARLIS/NA)
- Doris Duke Foundation for Islamic Art
- Hong Kong Maritime Museum
- National September 11 Memorial Museum
- Sterling and Francine Clark Art Institute Library
Archive-It: Crawl

COLLECTION

• Seeds
  Any document with a url, including
  • Images
  • Audio
  • Video
  • PDFs
  • HTML

SCOPE

• Which URLs are captured and which are not

FREQUENCY

• How often seed is crawled
What is a seed?
A seed is any URL that you want to capture:

an entire website:
http://www.intenttodeceive.org/

a specific part of a website:
http://www.intenttodeceive.org/videos/

a specific URL:
http://www.intenttodeceive.org/forger-profiles/han-van-meegeren/
Archive-It: Manage

Metadata
- Dublin Core
- Collection, seed, and document level

Support
- Training sessions
- Partner support
- User community

Storage
- Archive-It hosts content and back-up on multiple servers
- Partner can request copies of data
NYARC Archive-It Collections

https://www.archive-it.org/organizations/484
Collecting and Making Meaning: Challenges in Collection Documentation

Ex Libris Primo Proof of Concept
Better access to important websites

Grogan & Company, Inc.

Title: Grogan & Company [electronic resource]: fine art auctioneers and appraisers.

Publisher/Date: Dedham: Grogan & Company
Publication history: Began in 1999

Connect to:
- Current site
- Archived site

Note: Title from caption (viewed on June 11, 2013).

Summary: Grogan & Company's website offers users the ability to submit bids for current auction sales, view catalogs of current and forthcoming auction sales, and see prices realized at recent auction sales. Founded in 1987, they are located in Dedham, Massachusetts. They specialize in fine art, antiques, silver, jewelry, and oriental rugs.

LC subject: Grogan & Company
Other subject: Art auctions -- Massachusetts -- Dedham
LC subject: Auction houses -- Massachusetts -- Dedham
LC subject: Auctioneers -- Massachusetts -- Dedham
Genre/Form: Websites
WorldCat no. 854643471

Permanent link to this page: http://arcade.nyarc.org:80/record=b1213752~S16
Use these tools: They are free!

**INTERNET ARCHIVE**

**WayBack Machine**

Tools

- Wayback Machine Availability API
  Build your own tools.
- WordPress Broken Link Checker
  Banish broken links from your blog.
- 404 Handler for Webmasters
  Help users get where they were going.

**Save Page Now**

Capture a web page as it appears now for use as a trusted citation in the future.
Only available for sites that allow crawlers.
COLLABORATION WILL BE REQUIRED FOR THE FUTURE OF ART RESEARCH

- Art libraries
- Academic libraries
- Governments
- Publishers
- Scholars
- Tech companies
- Website and New Media creators
- Digital Humanities labs
- Professional organizations
Many Opportunities for Partnerships

- Build collections
- Influence systems development
- Extend our expertise
- Rethink our processes
- Sustain our mission
Links to resources cited, and other useful information on born digital content

http://www.dailydot.com/opinion/art-history-failing-internet/

Collecting and preserving the World Wide Web: a feasibility study undertaken for the JISC and Wellcome Trust (February 2003) http://www.jisc.ac.uk/media/documents/programmes/preservation/archiving_feasibility.pdf


Archiving the Web: A Case Study from the University of Victoria, Code(4)Lib Journal, Issue 26, 2014-10-21 http://journal.code4lib.org/articles/10015

http://webarchivingrt.wordpress.com/2013/12/06/it-is-time-to-embrace-the-present-guest-post-by-deborah-kempe/

http://nyarc.org/content/reframing-collections-digital-age, blog posting by Stephen Bury, June 18, 2012

New York Art Resources Consortium (NYARC) FAQs on Web Archiving http://www.nyarc.org/content/faq-web-archiving

Columbia University Web Archiving Summit 2012 https://webarch.cul.columbia.edu/


Archiving the Web for Scholars, by Steve Kolowich http://www.insidehighered.com/news/2011/05/06/libraries_try_to_preserve_and_archive_websites_for_academic_study

Overview of Web Archiving, by Jinfang Niu http://dl.org/dlib/march12/niu/03niu1.html


Further Resources


http://www.loc.gov/webarchiving/  Library of Congress Web Archiving

http://netpreserve.org/ website of the International Internet Preservation Consortium (IIPC)

SAA Web Archiving Roundtable http://webarchivingrt.wordpress.com/

Archive-It Knowledge Center https://webarchive.jira.com/wiki/display/ARIH/Welcome

Guidelines for Preservable Websites / Columbia University Libraries
https://library.columbia.edu/bts/web_resources_collection/guidelines_for_preservable_websites .html

Supporting the Changing Research Practices of Art Historians / Matthew P. Long, Roger C. Schonfeld, April 30, 2014
OBRIGADO!