ARTIST FILES SPECIAL INTEREST GROUP MEETING

ARLIS/NA Annual Conference, Pasadena 2013
Friday, April 26, 4:45-5:45
Room 214

Artist Files SIG Co-Coordinators
Sally McKay, The Getty Research Institute, SMckay@getty.edu
Samantha Deutch, The Frick Art Reference Library, deutch@frick.org

<u>Presenters:</u>
Kristine Hanna, Archive-it , kristine@archive.org
Anne Simmons, The National Gallery of Art, A-Simmons@NGA.GOV

<u>In Attendance:</u>
Arwen Spinosa (The Ringling), Billy Chi-Hing-Kwan (Philadelphia Museum of Art), Clayton Kirking (New York Public Library), Debra Roger (Portland Art Museum), Faith Pleasanton (The Metropolitan Museum of Art), Inge Reist (The Frick Art Reference Library), Jeff Snapp (Los Angeles County Museum of Art), John Evans (MFA Houston), Julia Kim (Los Angeles County Museum of Art), Kraig Binkowski (Yale Center for British Art), Lindy Narver (Laguna Beach Art Museum), Megan Oliver (The Ringling), Stephen Bury, The Frick Art Reference Library, and Traci Timmons (Seattle Art Museum)

<u>Recorder:</u>
Samantha Deutch

<u>Welcome</u>
<u>Updates</u> – Sally and Samantha gave updates on the WIKI database and background information on the Archive-it Pilot Project.
    WIKI

- Database no longer corrupted
- Will be able to add entries without extensive training
- Will be able to add artist names to Directory

    Archive it Pilot Project

- Recap from last year's Archive-it Workshop
- Background on Archive-it Pilot Project

<u>Kristine Hanna Archive-it Presentation</u>
–Kristine defined web archiving, different ways their partners have used archive it, why it is important, explained how Archive-it (**www.archive-it.org**) is used for archiving the internet and how the archived information can be accessed, and finally some challenges.

– What is web archiving
-    o   Web archiving is the process of collecting portions of web content, preserving the collections, and then providing access to the archives - for use and re use.

- What is a web archive
  - 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.
- Examples of how and why people are archiving the net
  - Collecting by theme, institutional archive (institutional memory/history), research (augmenting collections), online ephemera (virtual vertical files).
- Why is it important: Limited lifespan of a webpage
  - Misconception that content that exists on the web will remain there forever.
    - A report in Scientific American claims 44 days.
    - A subsequent academic study in IEEE suggests 75 days.
    - A Washington Post article indicates the number is 100 days.
    - Over 95% of government information today is born-digital. But less than 50% is being maintained with an active preservation plan. State of the Federal Web Report
- Archive-it as an archiving tool
  - Web based application allowing users to create, manage, access and store collections of digital content
  - Fully hosted solution—includes access and storage.
  - Provides tools for selection and scoping including cataloging with metadata
  - Ability to capture content using 10 different crawl frequencies
  - Can browse archived content 24 hours after a capture is complete; and full text search is available within 7 days
  - Restricted access options are available
- Challenges
  - Selection – what to capture and how much to capture
  - Metadata– Standards, levels (collection, webpage), scraped metadata (from the html of web pages)
  - Access—Public or Private
  - Technical—Dynamic content (Flash and JavaScript), new and mixed technologies (some old and outdated, others at the cutting edge).

Anne Simmons Presentation

- "Online Art Ephemera: Web Archiving at the National Museum of Women in the Arts," by Heather Slania, *Art Documentation: Journal of the Art Libraries Society of North America*, Vol. 32, No. 1 (Spring 2013), pp. 112-126 (http://www.jstor.org/stable/10.1086/669993)

- Presented the AF SIG pilot project with Archive-It with Anne Simmons (NGA) and Heather Slania (NMWA).
  - The project took on a hybrid approach and focused on capturing information on 50 contemporary artists and capturing information from 2 contemporary galleries (303 Gallery, http://www.303gallery.com/ , and Marianne Boesky Gallery, http://www.marianneboeskygallery.com/)
  - Working through the use of Google docs and e-mail Anne and Heather compiled a list of URL seeds for Archive-it to crawl and capture information

- Anne showed her first report which illustrated some content capture hurdles that needed to be overcome. Most of the issues were resolved by ignoring robots, but some of the issues, such as lost headers could not be controlled as the dynamic content, specifically flash, could not always be captured.
- Limiting the crawl
  - Careful curation was needed in order to capture enough useful information but not adding information that leads in different directions.
- Anne illustrated the reports and showed examples from the captured content
  - The Artist Files SIG collection is publicly available (http://archive-it.org/organizations/692;jsessionid=6E61856E333B363B4772B8555F5BBD5D)
- After resolving some of the capture difficulties by ignoring robots Anne and Heather focused on adding metadata.
  - They focused on two fields: creator and subject, employing LC Headings.
  - They added metadata by document or URL and this can be completed by adding information as a batch or individually.

- What type/format of information was captured?
  - Html
  - Multiple image formats
  - Flash
  - Video
- Some Questions posed by Anne and Heather:

  - Granularity
  - What types of material to capture? Obviously artist websites and pages from galleries. What about Facebook pages? Blogs? Webpages from major museums?
  - Workflow, collaboration
  - Access: how does this resource fit into the world of online art information?

Project Follow-up in Six Months:
The collection is live and we'll run another crawl in 6 months to see how the captured content has changed

Discussion:
How would we scale up the project?
How can you capture information from local newspapers?
How to handle permissions? Information is publically available can it be crawled and saved for future research?
Technical issues: More and more dynamic websites –RSS, Video, Images, Flash, Dark deep space
Multiple formats new Challenges
How to get buy in from the producers of content (Artist, Galleries, Auction Houses, etc. . .)?
Could we form a consortium of sorts?
How to keep repetition from occurring?

What is needed to make web-archiving successful?

How can we make more progress and establish a "vision"?

What features might the community want to have the Internet Archive help?

Meeting Adjourned.