What Are You Going to Do With that Data?: Digital Collections and Humanities Research

Harriett Green
English and Digital Humanities Librarian
University of Illinois at Urbana-Champaign

KU Digital Humanities Forum 2012

September 22, 2012

Digital Collections: A Bounty

PROJECT bamboo

http://www.projectbamboo.org

Mission:

"How can we advance arts and humanities research through the development of shared technology services?"
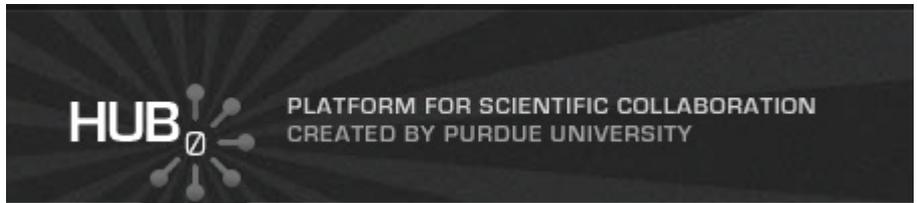
Partner Institutions:

- Australian National University
- Indiana University
- Northwestern University
- Tufts University
- University of California, Berkeley
- University of Chicago
- University of Illinois, Urbana-Champaign
- University of Maryland
- University of Oxford
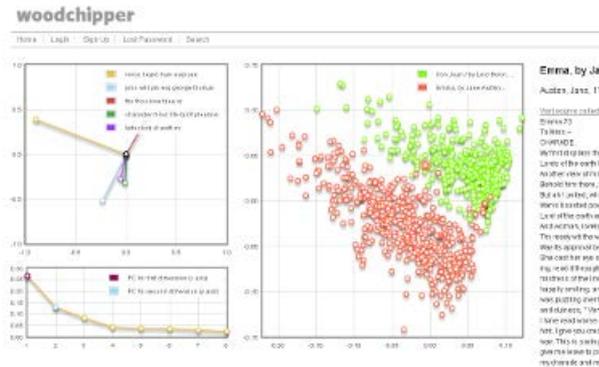- University of Wisconsin, Madison

PROJECT bamboo

http://www.projectbamboo.org

SEASR and Meandre

Woodchipper

HUB Zero

Oxford Virtual Research Environment for the Humanities

Scholars and Digital Collections Study

- Why and how do humanities scholars use digital collections in their research?
- What do scholars need in the functionalities and features of digital collections in order for them to be useful in humanities research?

- Study of faculty: Survey and interviews

Participating Institutions

- Indiana University
- Michigan State University
- Northwestern University
- Penn State University
- University of Chicago
- University of Illinois at Chicago

- University of Illinois at Urbana-Champaign
- University of Iowa
- University of Maryland
- University of Minnesota
- University of Nebraska at Lincoln
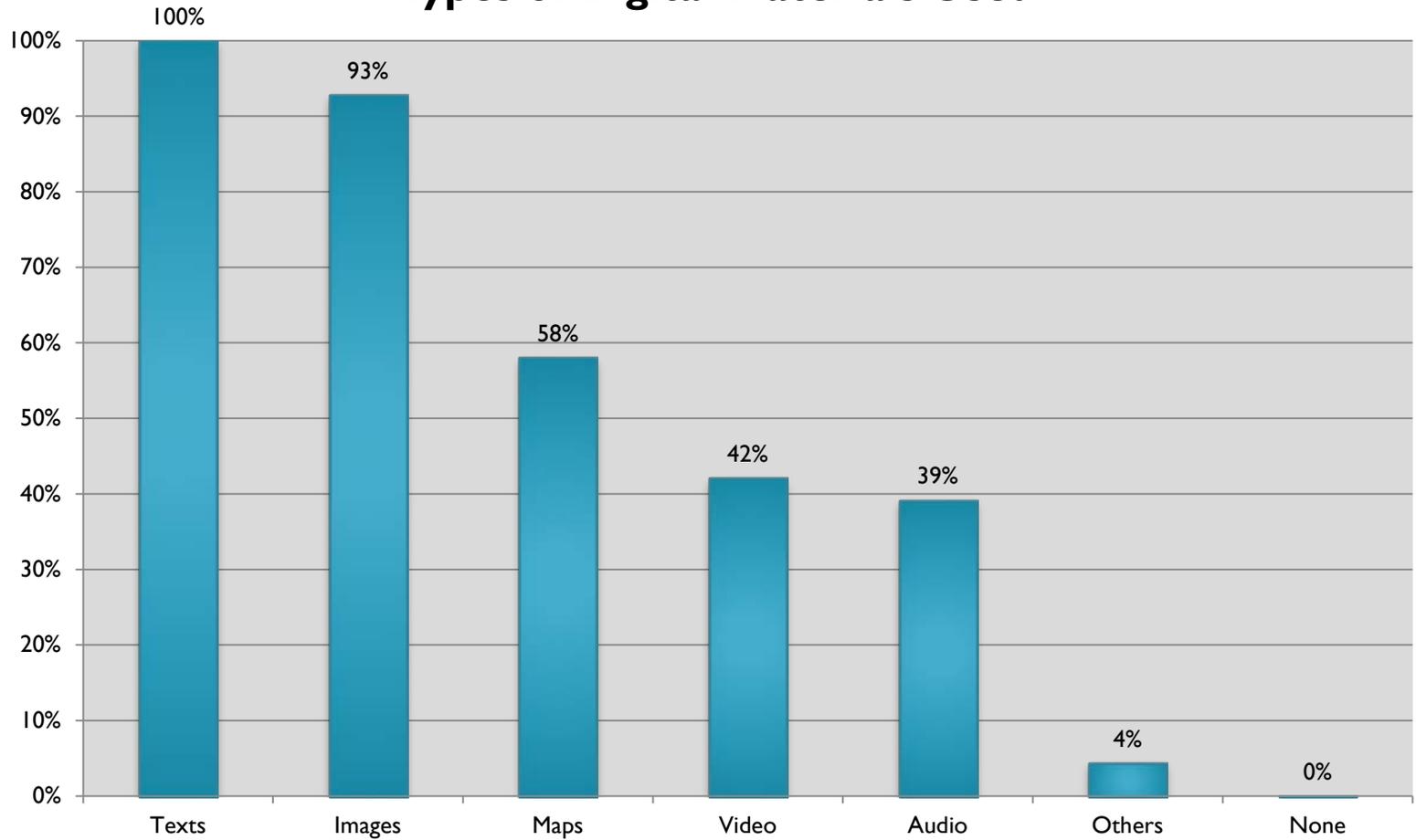- University of Wisconsin–Madison

Study Components

SURVEY

- Web-based multiple choice and text entry survey

- Distributed to randomly selected one-third of English and History faculty members at each institution

Types of Materials

Types of Digital Materials Used

Bar chart showing percentages: Texts 100%, Images 93%, Maps 58%, Video 42%, Audio 39%, Others 4%, None 0%.

Frequency of Use

Frequency of Use of Digital Materials

Legend:
- Never
- Less than half
- Equal for print & digital mat.
- More than half
- Always
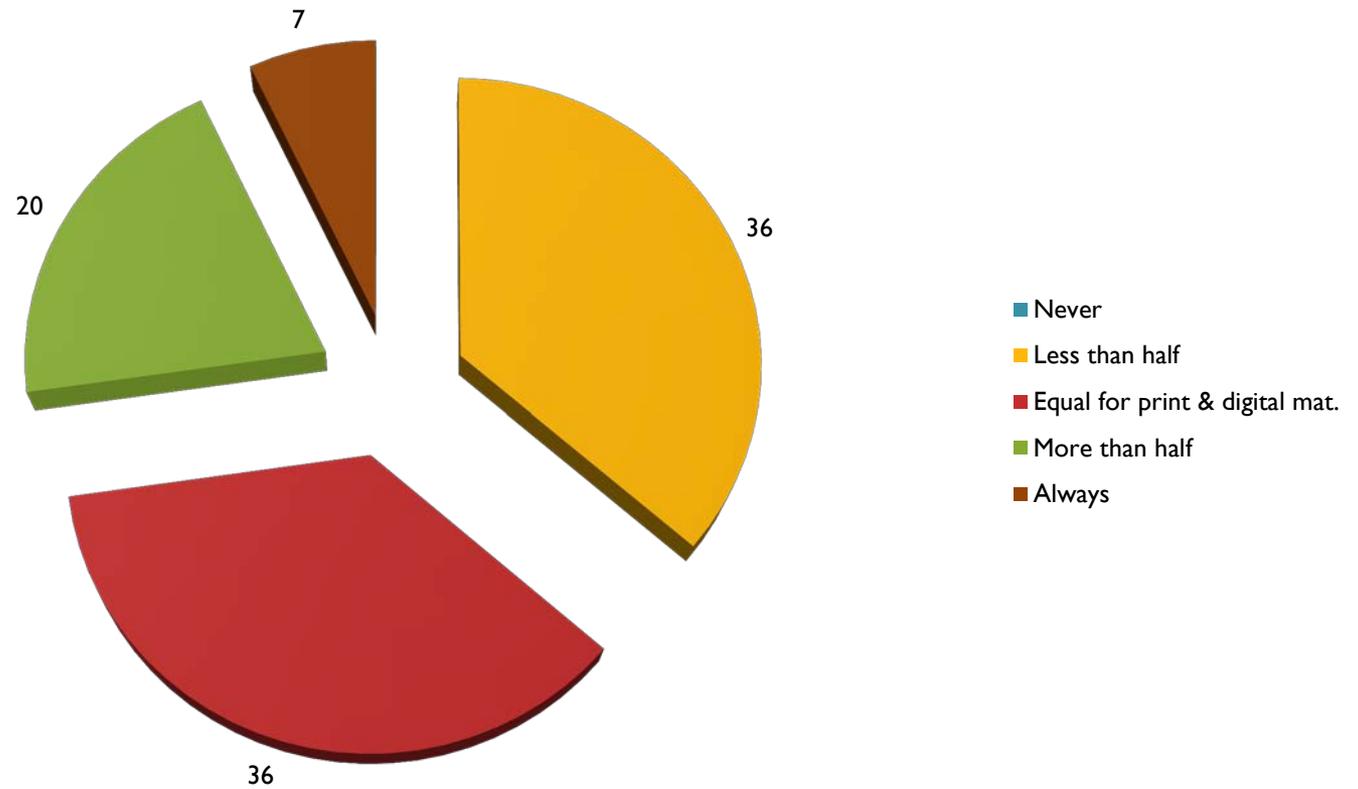
Text Collection Preferences

- "links to related material or bibliographic resources"
- "metadata must include rigorous bibliographic information and stated principles, either explicitly derived from existing modern critical editions or stated as unique to this database."
- "ease of access to items within collection (not half a dozen mouse clicks to get to it)"
- "capability to be read on multiple devices (e.g., iThing, laptop, desktop)."
- "The ability to annotate these texts in private and public environments, where I could keep stuff for myself or share it with others."
- "Use of JavaScript or some system that is compatible with software loaded on most research faculty's computers (I have colleagues who often work on out of date computer systems at home)"

Image collection preferences

- "Images must cooperate with both Powerpoint and Word."
- "Ability to export in multiple formats (i.e. jpegs and higher resolution tifs)"
- "information on permissions and how to request them"
- "ability to have a folder of images in the collection to which I can return that contains all of the copyright and original publication material associated with the image"
- "viewing and zooming tools, color accuracy, ability to export, reliable metadata"

Multi-media collections preferences

- "video ability to moderate sound and size"
- "ability to embed in other sites (I use my own websites for drafts of work in multi-media)"
- "ability to export and listen/view in another session"
- "Annotation tools"
- "detailed metadata, a time counter for audio and video, downloadable"

Possible To Do

- ◦ "exporting files and creating my own text and visual files either for teaching or research purpose"
- "ability to create one's own "library" held on the digital database's server space (like Hathi Trust collections allow) -- a concern for those of us who don't want to eat up all our hard drive space with PDFs.
- "knowing what is available! This is not obvious!"
- "high resolution (I do a lot of on-line work and my eyes are really degenerating)"
- "More materials—access to obscure or rare materials."
- "Ability to know copyright issues or costs up front."

How do you use digital collections?

- "I use specific sites to see what might be at a collection that is not available online and may require a visit--like the Southern Historical Collection or the Newberry Library."

- "I use databases that offer digital collections of all ancient Greek texts ever written to search for specific words and all the texts these words have ever appeared in. . . . I have used digital collections of images of inscriptions to check whether the text is correct or to amend the text.

- "I frequently need to use Japanese language texts not available in our university library (some not available in the U.S. at all). These materials can be found in digital archives maintained overseas."

Interviews

- Follow-up with volunteers from survey respondents
- Faculty from performing arts, art history, interdisciplinary humanities centers, or digital scholarship centers at each institution

Interviews

- Use of digital materials versus physical materials

"I use the "real" objects if possible. I use digital reproductions if I don't have access to the originals. Since I don't have access to many of the original objects and images I want to study, I would guess about half and half."

Interviews

- Benefits of Original
  - Completeness of contents
  - Sensory experience
  - Integrate print and digital

"I still must and do consult originals, but for teaching I have always had to rely on surrogates, and for my research, the availability of databases particularly for periodicals, historical dictionaries, and other online primary and secondary sources generally has been an enormous boon."

Interviews

- Benefits of Digital
  - Portability
  - Wide accessibility
  - Preservation

  "I teach contemporary literature so no one expects me to assign a given writer's drafts of a published novel. Why should it be any different when it comes to digital images of art that may be housed in distant museums or digitized manuscripts that may be too frail to sustain the wear and tear of hundreds of freshmen's hands?"

Interviews

- Areas to Improve
  - Searchability
  - More content
  - Annotation and editing tools: "The easier objects are to repurpose, remix, and reuse the better."
- Current challenges
  - Poor interfaces and user design
  - Incomplete contents
  - **Too much data?**

The Future?

- "If I could create a personalized launch page where I can access my most frequently used journals or other electronic texts or images without having to go through a laborious login process." → **Humanities cyber-infrastructure**?

- "Enabling anything like seamless access to the cultural record will require developing tools to navigate among vast catalogs of born-digital and digitized materials, as well as the records of physical materials."—ACLS, *Our Cultural Commonwealth* report (2006)

Thank you!

Harriett E. Green

English and Digital Humanities Librarian

University of Illinois at Urbana-Champaign

green19@illinois.edu

Twitter: @greenharr