

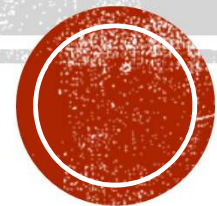
# DATA PRACTICES IN DIGITAL HISTORY

Rongqian Ma ([rom77@pitt.edu](mailto:rom77@pitt.edu))

Fanghui Xiao ([fax2@pitt.edu](mailto:fax2@pitt.edu))

School of Computing and Information, University of Pittsburgh

Feb. 5, 2019



The use of computational methods for humanities studies and rise of humanities computing



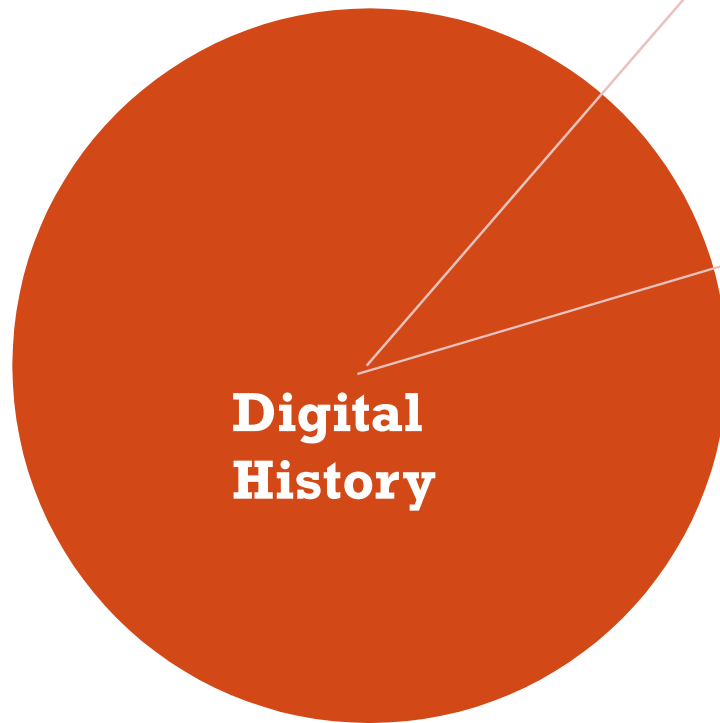
The term "digital history" first appeared in the name of "Virginia Center for Digital History" in 1997



First attempt to define "digital history" in 2008 on a discussion roundtable titled "Interchange: The Promise of Digital History"

## BRIEF HISTORY OF DIGITAL HISTORY

- **“Digital history** is an approach to examining and representing the past that works with the new communication technologies of the computer, the Internet network, and software systems. On one level, digital history is an open area of scholarly production and communication, encompassing the development of new course materials and scholarly data collections. On another, it is a methodological approach framed by the hypertextual power of these technologies to make, define, query, and annotate associations in the human record of the past. To do digital history, then, is to create a framework, an ontology, through the technology for people to experience, read, and follow an argument about a historical problem.” (Cohen et al., 2008, p. 454)



**For readers:** form their *own* understanding of the past by actively engaging with historical sources within a multimedia virtual environment and explore histories from a hypertextual environment

**For digital historians:** continual adding, annotating, editing, and refining of resources and narrative

*The Valley of the Shadows, Race  
and Place: An African American  
Community in the Jim Crow  
South, and etc*

**Argumentation**

**Future model of  
digital history**

Interdisciplinary collaboration with programmers,  
information architects, designers, and publishers

Tooling up historians to deepen  
historical analyses

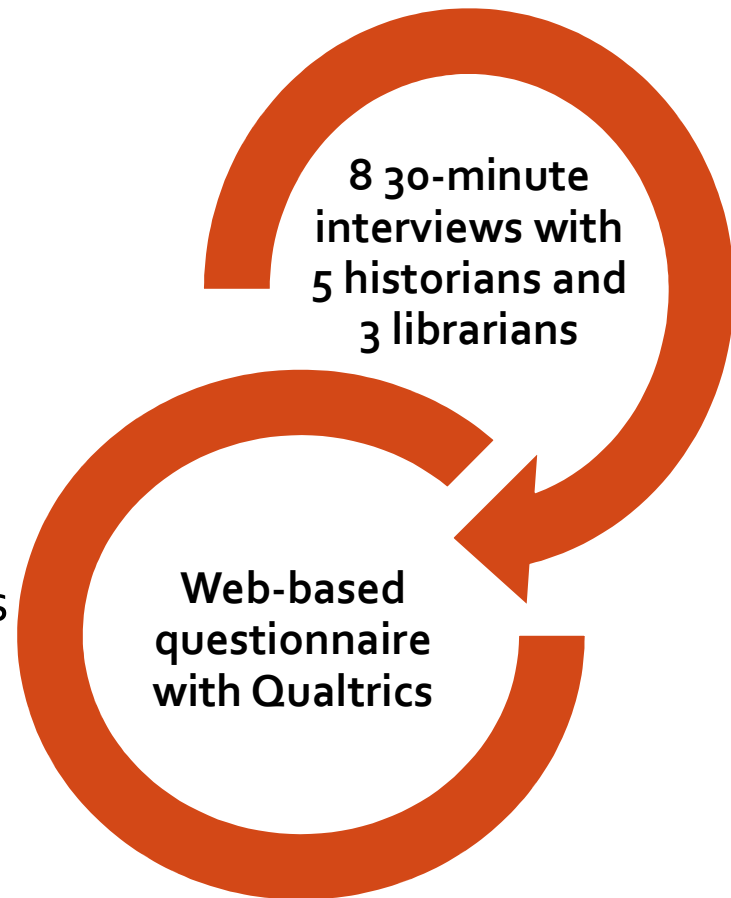
Shifted focus to data (e.g., data-driven  
approach to digital history, linked open data)

# RESEARCH QUESTIONS

- 1. From an ontological perspective, what does the notion of “data” and the digital approach add to historical research?
- 2. What are the current data practices in digital history research? How do data practices in digital history differ from conventions in analog history and add to the modes of scholarship production in the history discipline?
- 3. What are the major challenges that current digital history researchers face in terms of data practices?
- 4. Considering the current data practice issues and challenges in digital history, what can be done to improve data practices in digital history?

# METHODOLOGY

- ✓ Iterative sequential mix-methods design
- ✓ Grounded theory approach
  - ✓ Semi-structured interview
  - ✓ Web-based questionnaire distributed to three email lists and personal networks received 40 valid responses



Data practices: data collection, processing, preservation, analysis, presentation and sharing

Attitudes, thoughts, and concerns toward digital history

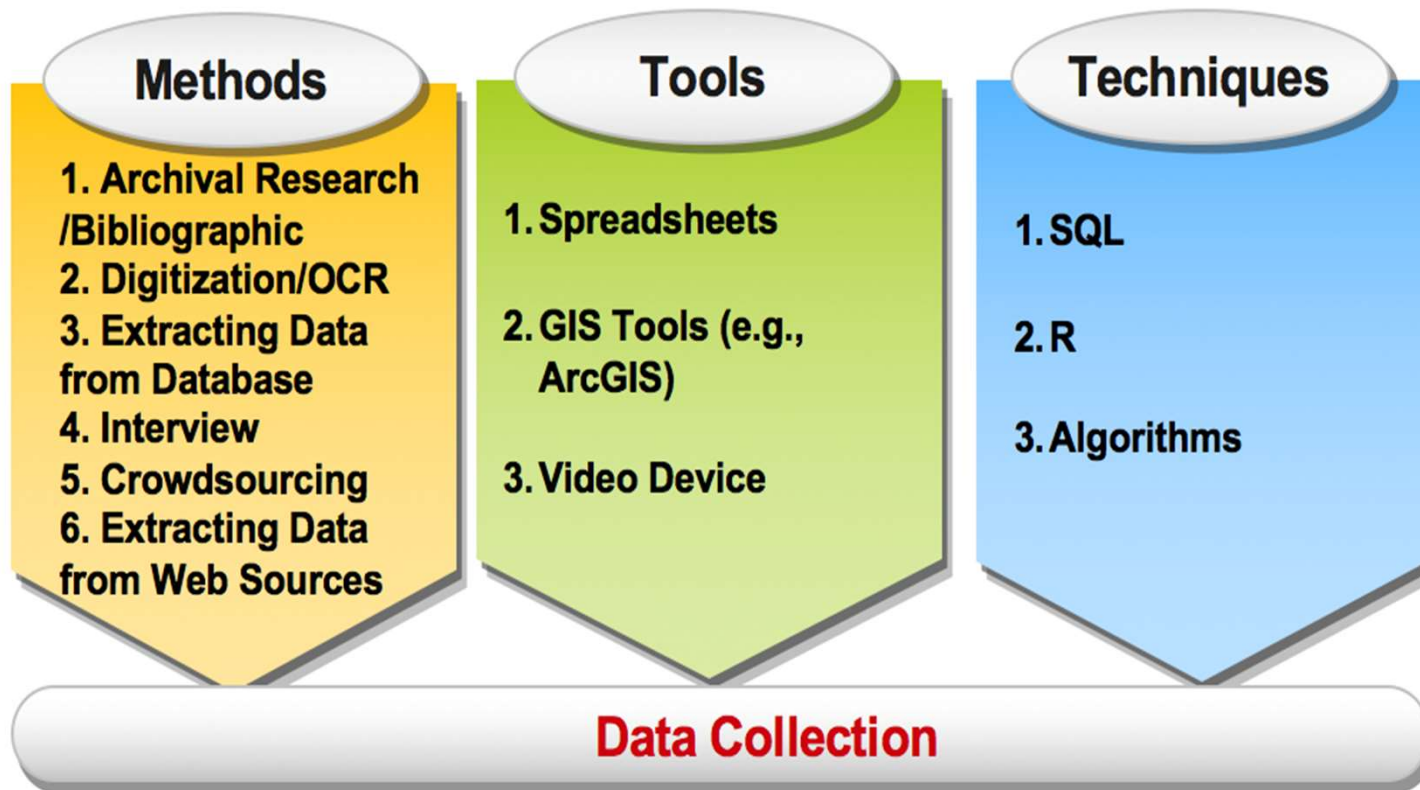
Collaboration practices

# THE “DIGITAL” AND HISTORY RESEARCH

- Perform research with larger datasets and share research outcomes more broadly
- Raise new research questions and identify new resources
- Integration of varied computational methods for analysis
- Raise importance of communication
- Better track the uncertain relationships between data and historical analysis
- No fundamental difference.

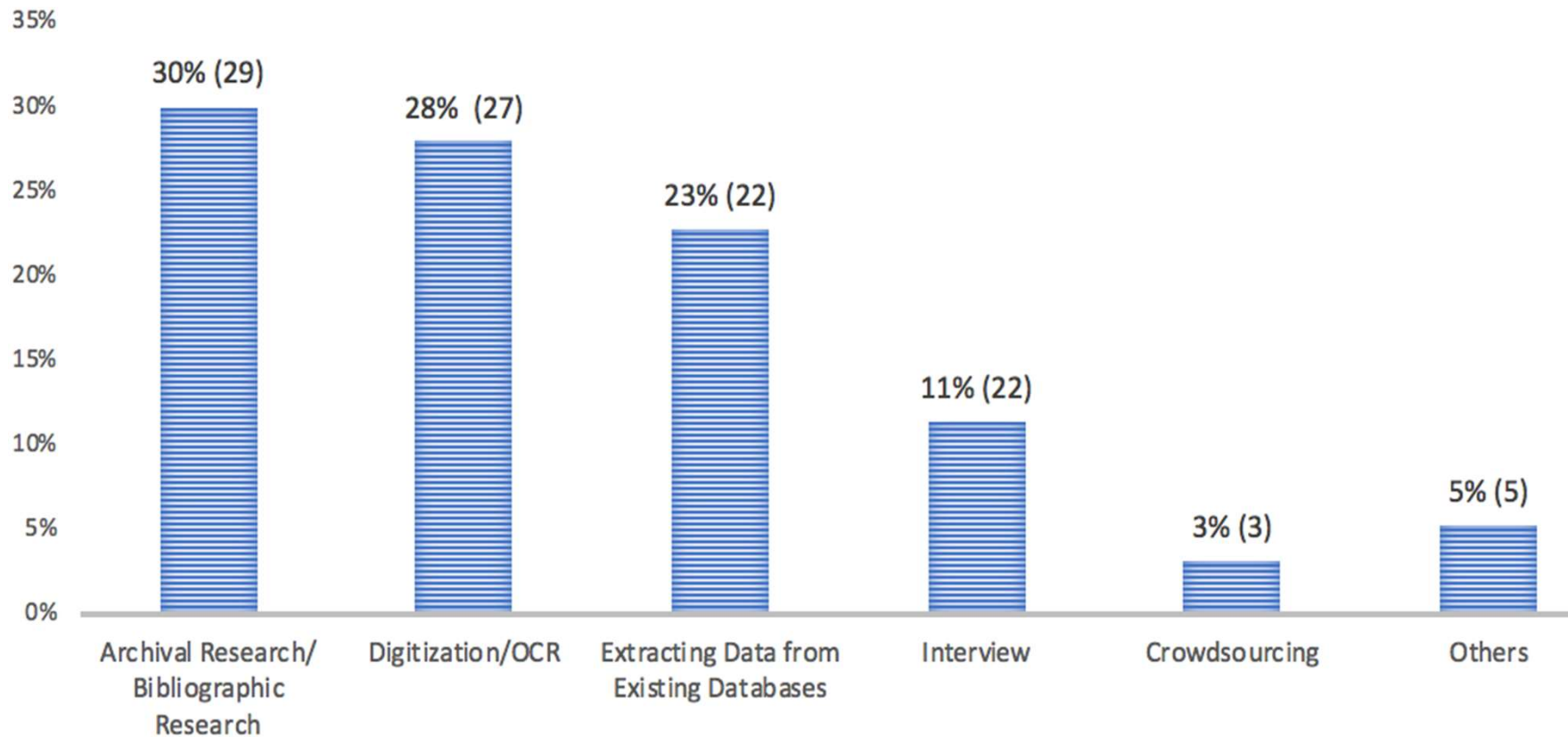
“I think the work of history is always trying to find materials that provide some kind of insight about the human past and how it changes over time,” said the interviewee, and “digital work is just part of that same process” (S2)

# DATA COLLECTION



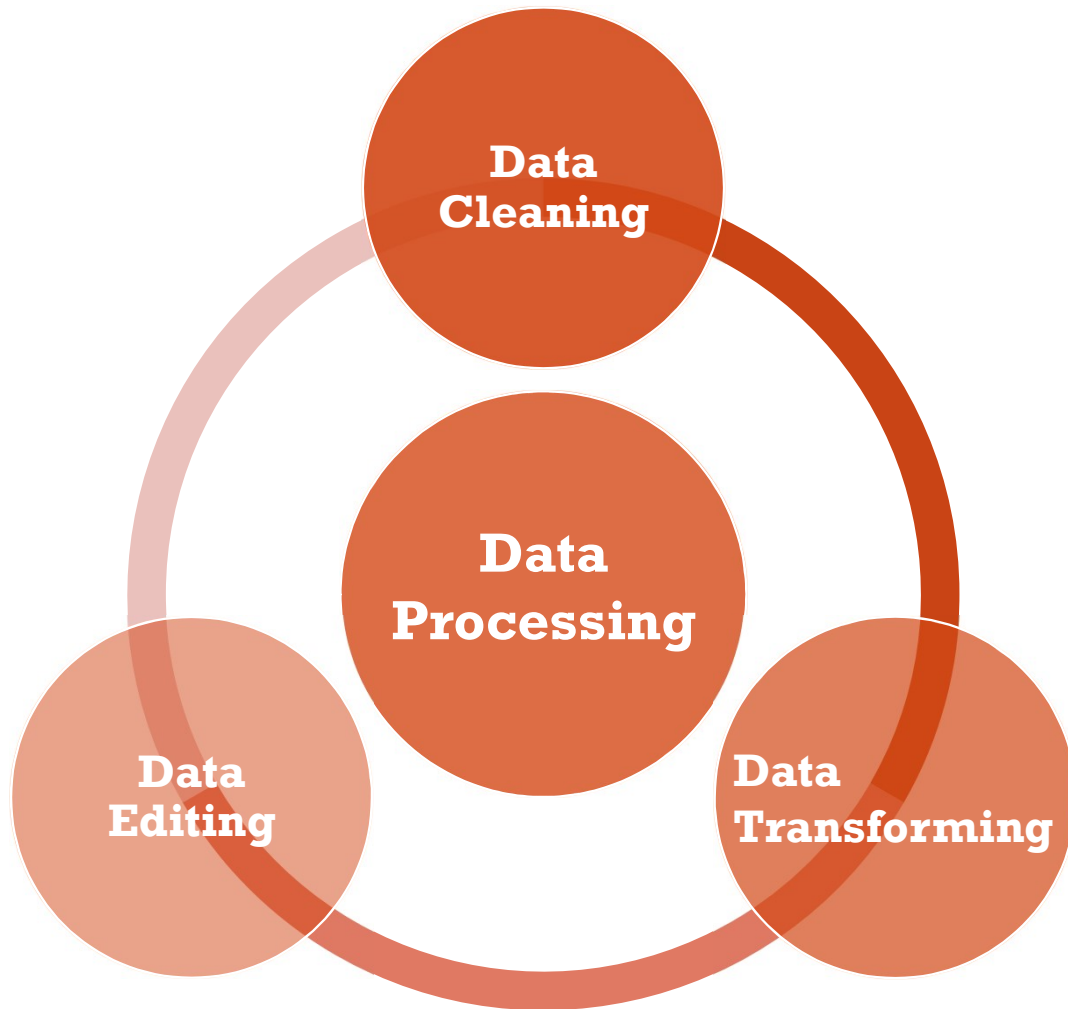
**Figure 1.** Data collection methods, tools, and techniques

# DATA COLLECTION



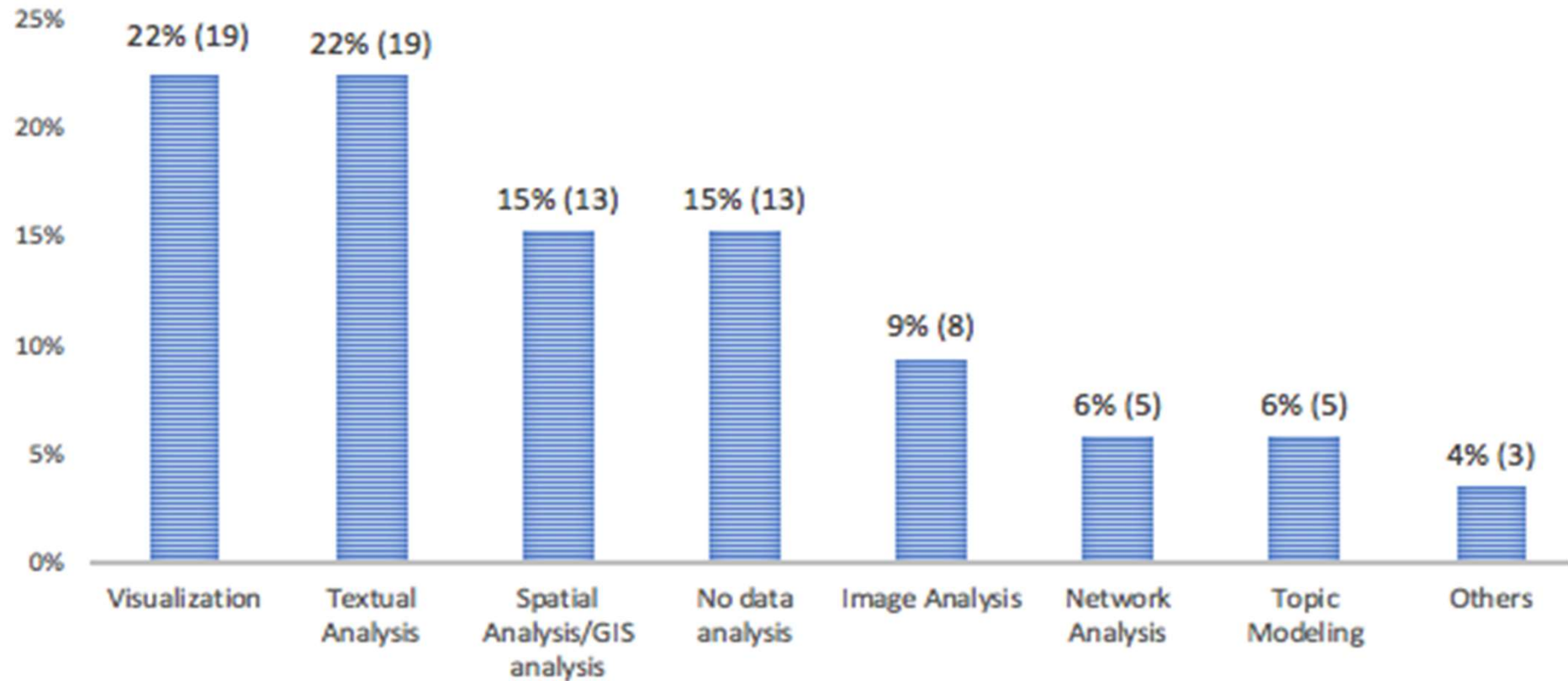
**Figure 2.** Distribution of data collection methods

# DATA PROCESSING



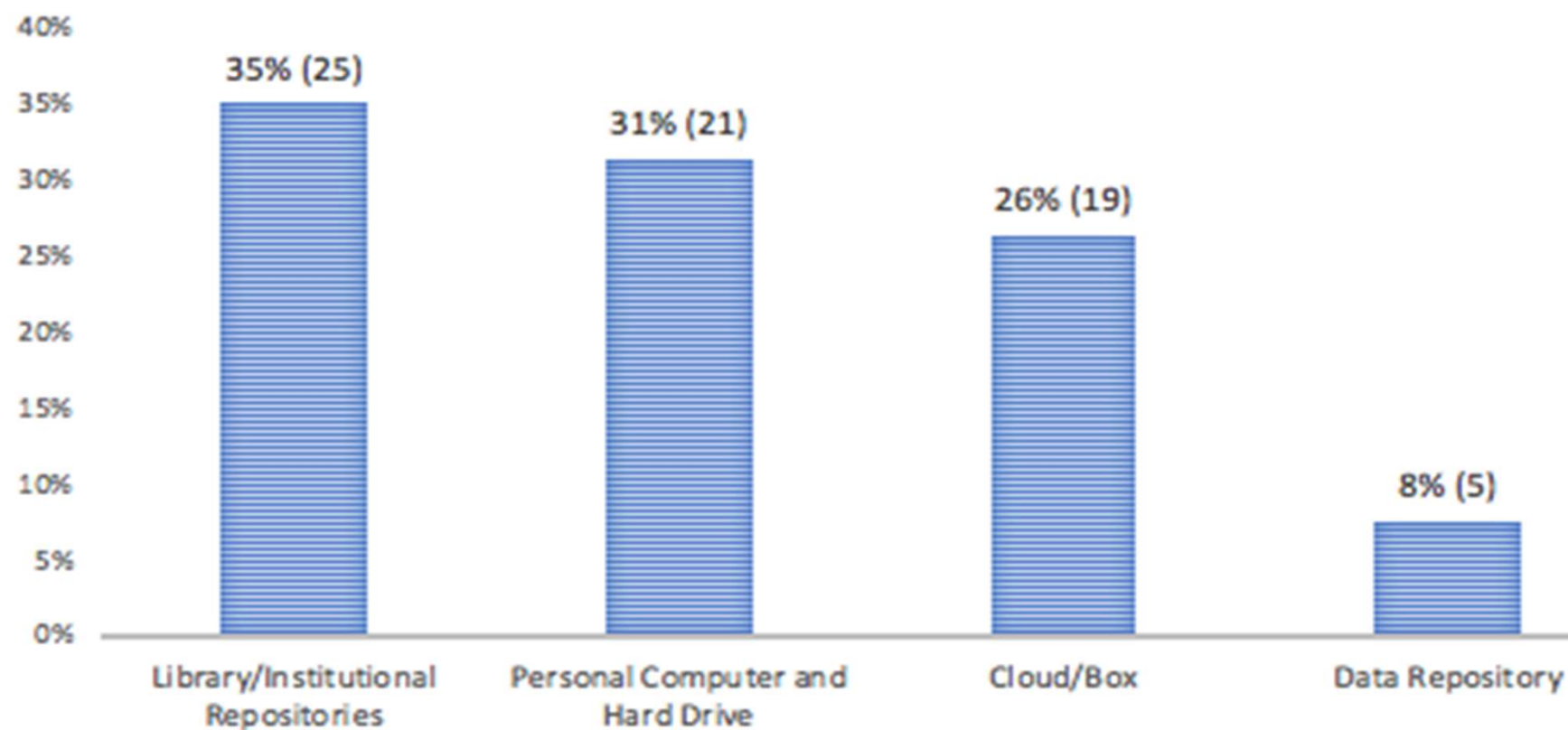
*"The vast majority of digital humanities projects, at least in history, is about cleaning data. You got the data, but you spent the majority of your time cleaning them, getting them into correct format to work with whatever software that you use for analysis or visualization." (S1)*

# DATA ANALYSIS



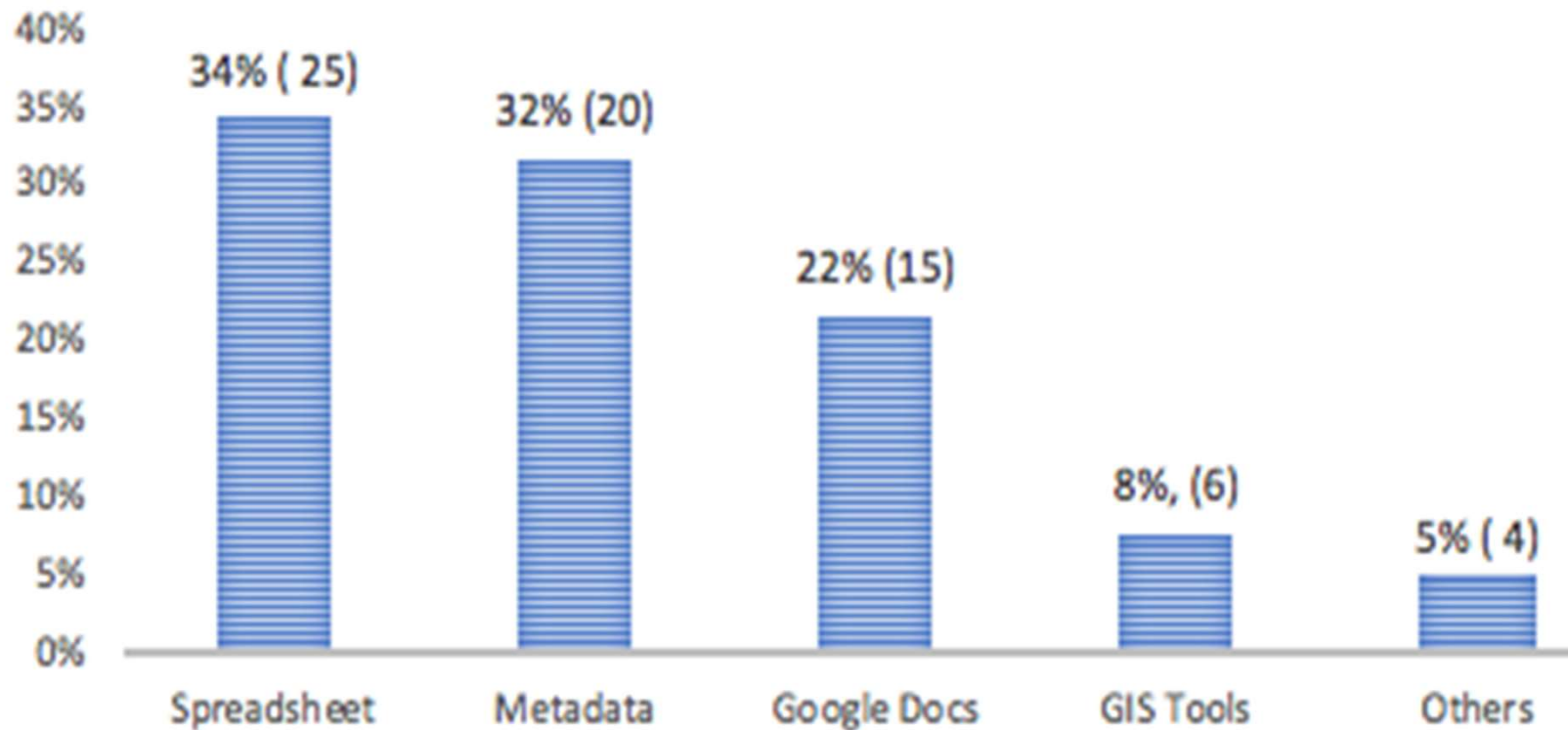
**Figure 3.** Distribution of data analysis methods

# DATA PRESERVATION



**Figure 4.** Distribution of preservation locations

# DATA PRESERVATION



**Figure 5.** Distribution of preservation tools or techniques

# DATA PRESENTATION AND SHARING

**Table 1.** Channels of data presentation and sharing

Channel	Number	Percentage
Website	23	18%
Digital Collection	22	17%
Conference paper and presentation	19	15%
Public presentation and lecture	16	12%
Journal article	14	11%
Interactive Visualization	9	7%
Dataset	8	6%
Database	8	6%
GitHub	3	2%
Monograph	2	2%
Dissertation	2	2%
Others	3	2%

# DATA CHALLENGES IN DIGITAL HISTORY

**Table 2.** Common challenges in digital history projects

<b>Challenges</b>	<b>Number</b>	<b>Percentage</b>
<b>Lack of technical skills</b>	21	23%
<b>Support (e.g., institutional, departmental, financial)</b>	17	19%
<b>Data quality issues</b>	14	16%
<b>Limitations in existing preservation platforms</b>	12	13%
<b>Difficulties of finding needed data</b>	10	11%
<b>Communication problems among team members</b>	6	7%
<b>Limitations in existing presentation platforms (e.g., interactive visualization)</b>	6	7%
<b>Others</b>	4	4%



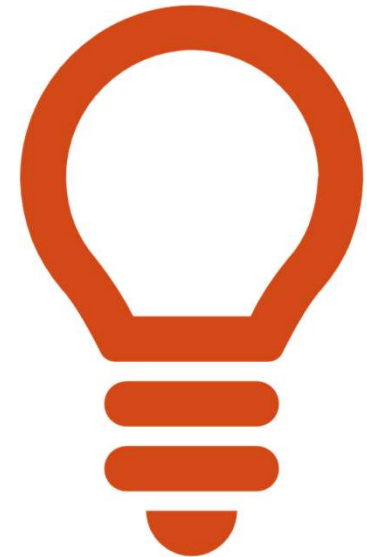


## DISCUSSION

- Data sharing
- Awareness of metadata
- Digital history training
- Transformation of the Roles of Libraries and Librarians in Digital History
- Deeper collaboration
- What can libraries do to improve data practices in digital history?

# WHAT CAN LIBRARIES OFFER?

- *Understanding* particular data needs and requirements for digital historical research for better engagement with digital history research
- Creating specifically-tailored *data practice guidelines* for historical research
- Providing *consulting* and *training* services (e.g., workshops)
- Constructing and promoting *data infrastructures* to support data practices in digital history
- *Leadership* in digital history; Being more proactive in reaching out to historians and other partners to address data challenges and developing leadership roles in digital history projects.



**THANK YOU!**

---

Q & A