# EVALUATING GITHUB AS A PLATFORM OF KNOWLEDGE FOR THE HUMANITIES

**Lisa Spiro and Sean Morey Smith**

Rice University

*Digital Humanities 2016*

July 2016

# Exploring and Critiquing Platforms of Knowledge in the Humanities



**RICE** School of **Humanities** Humanities **Research Center**

**SAWYER SEMINAR 2015-16**

| Home | Info | Participants | Presentations | Workshops | Resources | Blog |

**Platforms of Knowledge in a Wide Web of Worlds: Production, Participation, and Politics**

*Mellon-funded*
*Sawyer Seminar on Platforms of Knowledge*

# Why Examine GitHub as a Platform?

- It seems to have relatively wide adoption in digital humanities.

- It facilitates the collaborative development of open source software.

- It raises questions about how and why we select platforms for digital humanities work.

- It offers insights into how digital humanists do their work.

# Methods

- **Identifying selected digital humanists** with GitHub (GH) accounts
  - Presenters at last three DH conferences who have public GH accounts
  - Staff affiliated with centerNet organizations who have public GH accounts
- **Surveying** selected DH GH users
  - who have email addresses associated with their accounts
  - who use English on their accounts (so they can opt out)
- **Interviewing** selected DH GH users
- **Analyzing public data** associated with selected DH GH user accounts (assuming they didn't opt out)

# PRELIMINARY RESULTS

# How many digital humanists have a GitHub account?

- 369 out of 1720 eligible presenters at DH 2013, 2014 and/or 2015 = ~**21%**
- Caveats:
  - Not everyone who attends DH identifies as a digital humanist, and not all digital humanists are at DH
  - Some repositories may focus on areas other than DH
  - False negatives are likely, especially if someone has a common name or doesn't associate their real name with a public GitHub account
  - A few false positives are also possible due to the challenges of name disambiguation
  - Having a GitHub account doesn't mean that one is an active or current user.

# Study Population

- Survey Respondents
  - 59 respondents, for a response rate of approximately 30%
  - Mostly male (81%)
  - Mostly North American (61%)
- Interviewees
  - 17 interviewees (also invited to take the survey)
  - Mostly male (88%)
  - Mostly North American (65%)

# Survey: What is your position?

| Answer | % |
|---|---|
| developer | 38.60% |
| professor | 36.84% |
| research staff | 22.81% |
| graduate student | 19.30% |
| other (please specify) | 15.79% |
| information technologist | 8.77% |
| librarian | 5.26% |

# Survey: Do you consider yourself a digital humanist?

| Answer | % |
|--------|------|
| Yes | 68.42% |
| Maybe | 26.32% |
| No | 5.26% |

# Survey: For what do you use GitHub?

| Answer | % |
| --- | ---: |
| Software development | 89.47% |
| Documentation | 54.39% |
| Datasets | 47.37% |
| Websites | 45.61% |
| Writing projects, e.g. books and articles | 40.35% |
| Course materials | 21.05% |
| Collections of online resources | 17.54% |
| Research notes | 14.04% |
| Syllabi | 8.77% |
| Other (please specify) | 1.75% |

# Interviews: Criteria for Evaluating Platforms

- **Usability:** "Are the interfaces reasonably intuitive?" (Jean Bauer)

- **Community: "**Find out who else it is using it and how they are using it." (James Baker)

- **User needs**: "Does it have the features and requirements that you need?" (Bauer)

- **Data portability:** "Can you escape and what will escape cost you[?]" (Eric Kansa)

- **Openness: "**Is it open source and connected to the open source community?" (Bauer)

# Survey: Why did you start using GitHub? (Free Text)



**Top Answers**

1. Version control
2. Collaboration
3. Sharing
4. Strong community
5. To be open
6. Used by home organization

# Survey: What are GitHub's strengths? (Check all that apply.)
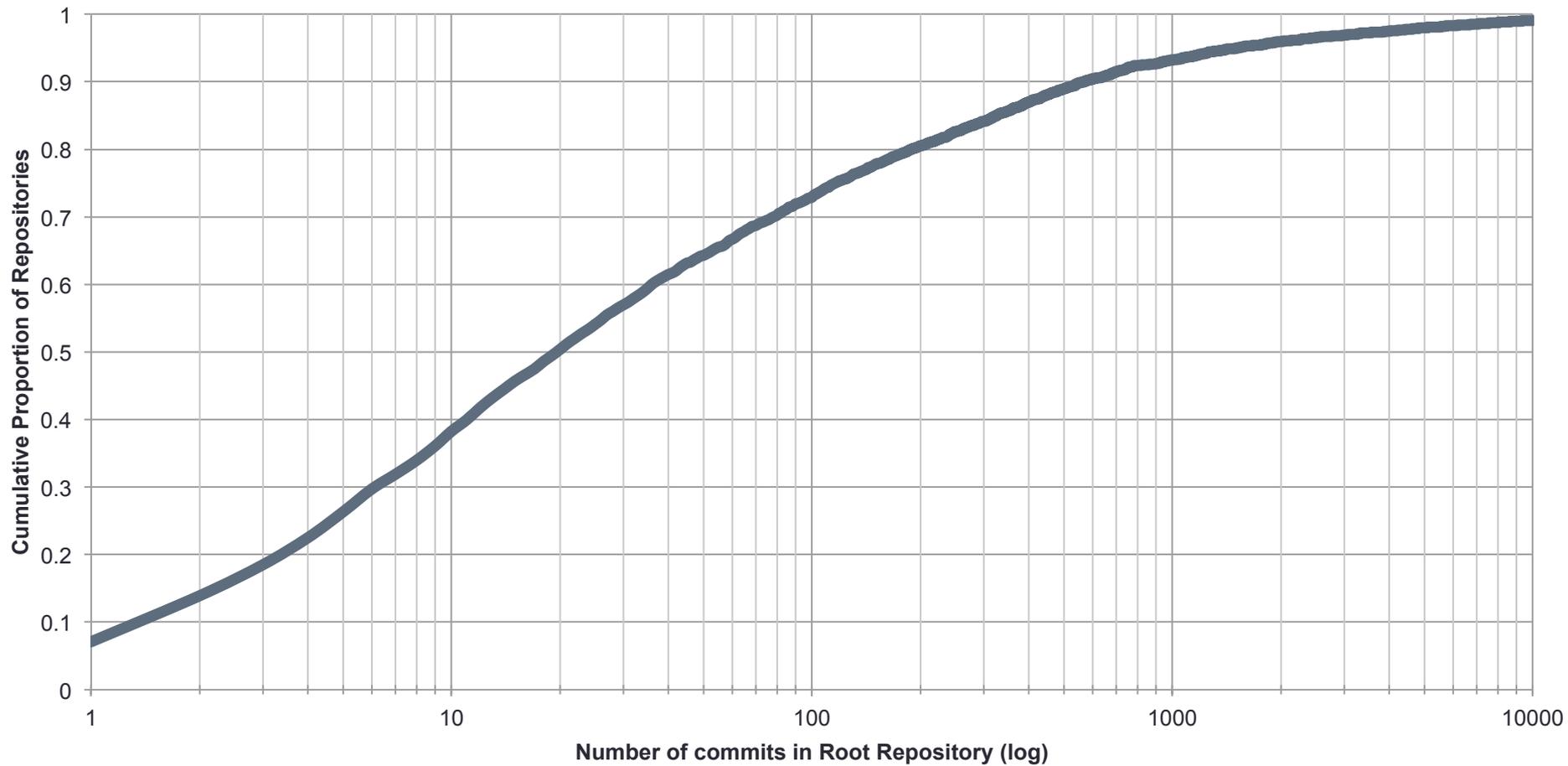
| Answer | % |
|---|---|
| Its version control features. | 98.25% |
| It provides free public repositories. | 94.74% |
| Its collaboration features. | 84.21% |
| It supports the open source community. | 75.44% |
| It enables reuse of code. | 66.67% |
| It is widely used. | 64.91% |
| My collaborators use it. | 56.14% |
| Its distributed architecture. | 38.60% |
| It allows GitHub users to connect with others. | 28.07% |
| It enhances the visibility of GitHub users. | 22.81% |
| Other (please specify) | 8.77% |

# Interview Results: Version Control

- Store and backup files on multiple computers
- Monitor changes efficiently; track bugs
- Recover from errors: "the ability to rewind time to one minute before disaster and restore" (Stephen Ross)
- Use different branches for production, development, experimentation
- Share and manage files among users: "see who committed what when" (Ross)

# Usage Data: Version Control

**Cumulative Ratio of Repositories with a Given Number of Commits**

# Interview Results: Collaboration

- Collaboration may involve
  - fellow members of a research group/ organization
  - the open source software development community
- How GitHub has supported collaboration
  - Using others' code—and learning from it
  - Providing feedback via issues
  - Coordinating contributions to open source projects
    - Potential for "democratization of scholarly editing" (Jonathan Reeve)
  - Working together on code or writing
    - Managing changes to text more efficiently

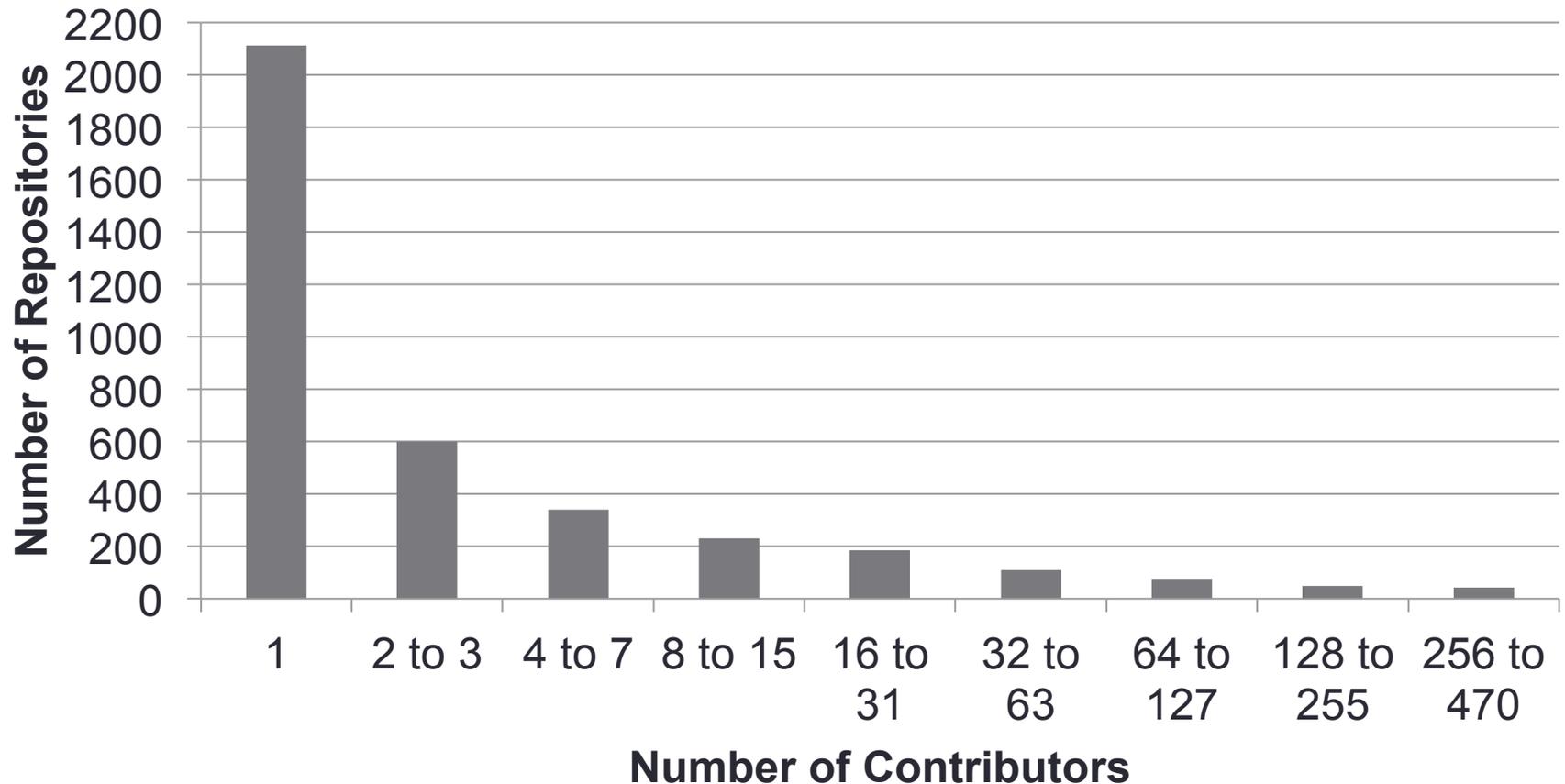# GitHub's Support for Collaborative Open Source Software Development

"The fact that almost all open source coders use GitHub in some way means that **it is trivially simple to find the software that you are using, make a change, improve it, and contribute it back**." (Jonathan Reeve, emphasis added)

# Challenges to Using GitHub for Collaboration

"You already have to have a community in place and/or be very pro-active about advertising and getting the project out there." (Tonya Howe)

# Usage Data: Collaboration

# Survey: What are GitHub's limitations? (check all that apply)

| Answer | % |
|---|---|
| It is difficult to get started with (learning curve). | 60.38% |
| It is owned by a for-profit company. | 32.08% |
| It is best suited for text, not other formats. | 28.30% |
| It does not provide for long-term preservation. | 22.64% |
| Other (please specify) | 22.64% |
| It is not very easy to use. | 18.87% |
| The complexity of its design | 11.32% |
| I am reluctant to make my work openly available. | 1.89% |
| Its terms of use and intellectual property policies. | 1.89% |
| I am concerned about my work being forked and reused. | 0.00% |
| Its social features. | 0.00% |

# Interview Results: Limitations of GitHub

- Difficulties learning & using Git & GitHub: "Git is hard… GitHub as a result suffers from that." (James Baker)

- GitHub does not handle data very well (especially large datasets)

- Some concern about relying on GitHub for long-term preservation, but also the sense that preservation isn't GitHub's purpose (& appreciation for libraries)

- Some uncertainty/ anxiety about GitHub's long-term business model

- A few concerns about inclusivity, most acutely a few years ago

# Future Work

- Complete data analysis, including further examination of gender and nationality.

- Explore perspectives of those who use GitHub very little or don't use it all.

- Compare usage of GitHub among digital humanists to the broader GitHub community.

- Crystallize recommendations for evaluating platforms.

# Thank You

- Interviewees
- Survey respondents
- John E. Sawyer Seminar on Platforms of Knowledge in a Wide Web of Worlds, sponsored by the Andrew W. Mellon Foundation
- Rice's Humanities Research Center
- Rice's Fondren Library
- Caleb McDaniel



Randall Munroe, xkcd (H/T Eric Kansa)