

# Use of GitHub as a Platform for Open Collaboration on Text Documents

Justin Longo  
University of Regina  
Johnson-Shoyama Graduate School of Public Policy  
3737 Wascana Parkway  
Regina, SK, Canada S4S 0A2  
306-585-5460  
Justin.Longo@uregina.ca

Tanya M. Kelley  
Arizona State University  
School of Public Affairs  
411 N. Central Avenue, Suite 400  
Phoenix AZ 85004-0687  
602-496-0457  
Tanya.M.Kelley@asu.edu

## ABSTRACT

Recently, researchers are paying attention to the use of the software development and code-hosting web service GitHub for other collaborative purposes, including a class of activity referred to as document, text, or prose collaboration. These alternative uses of GitHub as a platform for sharing non-code artifacts represent an important modification in the practice of open collaboration. We survey cases where GitHub has been used to facilitate collaboration on non-code outputs, identify its strengths and weaknesses when used in this mode, and propose conditions for successful collaborations on co-created text documents.

## Categories and Subject Descriptors

K.4.3 [Organizational Impacts]: Computer-supported collaborative work

## General Terms

Management, Performance

## Keywords

GitHub; Open Collaboration; Documents; Co-creation

## 1. INTRODUCTION

GitHub (<http://github.com>) is a software code-hosting web service principally used for software development, that augments the usability of the distributed version control system / source code management protocol Git by providing a web interface that automates some functions normally controlled through command line entries. GitHub also has social networking and project management features designed to enhance the capacity of users to work together, and interface features designed to lower the technical barriers to entry for new users [1].

As its user-base grows, the purposes for which GitHub is used have expanded to include all manner of digital products [5]. Recently, researchers and other observers are increasingly investigating the use of GitHub for purposes other than software coding and website development. Referred to as text, documents or prose, these non-code uses of GitHub reveal how the site

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owners/authors. *OpenSym '15*, August 19-21, 2015, San Francisco, CA, USA  
ACM 978-1-4503-3666-6/15/08.  
<http://dx.doi.org/10.1145/2788993.2789838>

provides a platform for social collaboration on non-code artifacts. As the uses and users of GitHub move beyond its core community of developers, the present and potential impact on fields such as social knowledge creation, open science, open collaboration and open governance warrants consideration of the conditions under which GitHub can facilitate collaboration in non-code domains.

Other open access platforms for collaboration certainly exist, including wikis, synchronous co-editing platforms like Google docs, and centralized file sharing repositories like SharePoint. However, GitHub includes unique features such as built-in social networking functions [6], back-end data capture and reporting [3], and principles of distributed version control and openness by virtue of the underlying Git architecture. GitHub allows for projects to be forked to accommodate alternative objectives and applications (subject to licensing), implements Git's distributed version control model using "pull requests" (PRs) to bring to the attention of the document owner proposed changes to the original, and uses cryptographic hash functions and "diff" displays to provide detail of changes made between versions.

While the use of GitHub for software development is being documented, its uses for other purposes are anecdotal, though growing [7]. We bring together observations from seven cases where GitHub has been used to facilitate collaboration amongst a number of co-contributors to non-code outputs. Our findings illustrate an evolving technological literacy and familiarity with GitHub, though also indicate that many barriers stand in the way of GitHub being used effectively as a platform for document collaboration. Modifications to the model will be required in order to improve its usability.

## 2. CASES IN OPEN DOCUMENT COLLABORATION

Open document collaboration cases were identified through a combination of searches of GitHub, looking specifically for mentions of document collaboration, projects we have contributed to or partnered with, and media reports and mentions in other literature of GitHub-based document collaboration efforts. We included cases where the dominant document formatting syntax used was plain text, a markup language (e.g., html) or GitHub's markdown format. While non-code contributions can also be made to any repository through the "Issues" function, we did not include this method in our sampling. Selected repositories were also deemed collaborative if they were open to accepting PRs from interested users without requiring some form of membership in the organization or prior permission to contribute to the project. The cases selected represent a range of academic, governmental, private sector and civil society initiatives.

We examined seven cases based on our selection criteria. One case was an organized effort that involved dozens of mathematicians completing a major book-length project.<sup>1</sup> Another math project had over 150 total contributors.<sup>2</sup> One politician seeking elected office in the United States made his platform available on GitHub and invited constituents to comment on and edit the documents.<sup>3</sup> Several examples exist of individuals attempting to generate collaborative efforts to write new legislation<sup>4</sup> or find improvements to existing legislation.<sup>5</sup> A magazine article that profiled the GitHub corporate culture was posted to GitHub itself and readers were invited to improve the article and add translations.<sup>6</sup> We also undertook our own experiment in open collaborative writing on GitHub by initiating an academic effort to co-create a literature review article.<sup>7</sup>

We reviewed the content of each of the cases examined and reviewed the process that produced that content. This was undertaken using the data inherent to the GitHub repository (i.e., number of contributors, commits per contributor, forks of the master repo, and issues and subsequent discussion), and supplemental descriptions such as blog posts and media reports.

### 3. FINDINGS

GitHub is purpose-built for collaboration around software, and with many alternatives available for document collaboration, attempts to undertake collaborative document writing in GitHub are rare. With its principal focus as a code-hosting and software development platform, we had difficulty finding many examples of true open collaboration on GitHub where the site was being used in more than an experimental way to post materials electronically, with passive consumption and few contributions from loosely-affiliated participants (e.g., cases 4 and 5).

For the new user, GitHub poses a very steep learning curve that limits contributions. It is a difficult platform for new, non-technical users to learn and is not well suited for text-based collaborations (e.g., case 7).

Stewardship of the collaboration process is vital, with guidance required from a core leadership team in order to maintain direction (e.g., cases 1, 2), an active contributor group to maintain momentum (e.g., cases 2 and 6), principles to guide participation and process (e.g., cases 1, 2, 3, 6 and 7), and a clear incentive structure in order to promote group sustainability (e.g., cases 1, 2 and 7).

GitHub allows for distributed workflows, with three main approaches flowing from that principle: centralized, integration manager, and Dictator and Lieutenants workflow; the choice has profound implications for the quality of the work and the volume and sustainability of contributions (e.g., cases 1, 2 and 7). Determining how to manage changes to a collaborative document is crucial in GitHub where asynchronous edits are made. Editors acting according to decision-making guidelines are needed to merge conflicts that occur in the process, as there are limited

automated ways of evaluating one text-based contribution against a conflicting one. Lessons from other environments such as wikis can be useful for evaluating the quality of contributions [2], whether through peer evaluation or group deliberation. Questions of how to coordinate group activity and multiple contributions must be addressed in seeking to regulate the work [4]. Finally, the topic under consideration must be conducive to both the process and the platform (e.g., cases 1, 2 and 6).

### 4. CONCLUSIONS

Our findings suggest that GitHub has some useful features for facilitating open collaboration on text documents and can be a useful tool when situated within a framework of guidance and rules-based interaction, but that the barriers to entry for non-technical users and its weaknesses when compared to other similar collaboration platforms limit its usefulness. We have responded to this by developing and deploying simplified tutorials to support new users. Layers such as prose.io can be used to increase usability, and the capacity to flag issues improves the ability of non-technical users to communicate ideas to project leaders. However, alternative existing platforms for document collaboration and the significant modifications required to make GitHub a more usable platform explain in part the limited number of document collaboration examples we found.

While that assessment describes its current state, this does not preclude the possibility of a significant rehabilitation of the GitHub platform toward one that is suitable for use by non-technical participants working collaboratively on documents. The strengths of GitHub – its openness, transparency, versioning and accountability – are the core of its value, and an ambitious goal would be to adapt the underlying GitHub architecture with a revised user experience more suited to document collaboration.

### 5. REFERENCES

- [1] Begel, A., Bosch, J., and Storey, M.-A. (2013). Social networking meets software development: Perspectives from github, msdn, stack exchange, and topcoder. *Software, IEEE* 30, 1, 52–66.
- [2] Biancani, S. (2014, August). Measuring the Quality of Edits to Wikipedia. In *Proceedings of The International Symposium on Open Collaboration*. ACM. OpenSym '14 , Aug 27-29 2014, Berlin, Germany 978-1-4503-3016-9/14/08.
- [3] Marlow, J., Dabbish, L., & Herbsleb, J. (2013, February). Impression formation in online peer production: activity traces and personal profiles in GitHub. *Proceedings of CSCW 2013* (pp. 117-128). ACM.
- [4] Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior*. In press. doi:10.1016/j.chb.2015.01.050
- [5] Storey, M.-A., Singer, L., Cleary, B., Figueira Filho, F., and Zagalsky, A. (2014). The (r) evolution of social media in software engineering. In *Proceedings of the Future of Software Engineering*, 100–116. ACM.
- [6] Tsay, J. T., Dabbish, L., & Herbsleb, J. (2012, February). Social media and success in open source projects. In *Proceedings of CSCW 2012* (pp. 223-226). ACM.
- [7] Zagalsky, A., Feliciano, J., Storey, M. A., Zhao, Y., & Wang, W. (2015, March). The Emergence of GitHub as a Collaborative Platform for Education. *Proceedings of CSCW 2015*, pp. 1906-1927.

<sup>1</sup> Case 1: <https://github.com/HoTT>

<sup>2</sup> Case 2: <https://github.com/stacks/stacks-project>

<sup>3</sup> Case 3: <https://github.com/coleforcongress>

<sup>4</sup> Case 4: <https://github.com/singpolyma/Copyright-Act-of-Canada>

<sup>5</sup> Case 5: <https://github.com/bundestag/gesetz#german-federal-laws-and-regulations>; <https://github.com/divegeek/utahcode>

<sup>6</sup> Case 6: <https://github.com/WiredEnterprise/Lord-of-the-Files>

<sup>7</sup> Case 7: <https://github.com/ASU-CPI/github-experiment>