12-12-2021

Emerging Collaboration and Topics in Open Climate Science: An Exploration of GitHub Repositories

Elizabeth Thompson
Purdue University, thompser@purdue.edu

Follow this and additional works at: https://aisel.aisnet.org/treos_icis2021

Emerging Collaboration and Topics in Open Climate Science

An Exploration of GitHub Repositories

Elizabeth R. Thompson, thompser@purdue.edu

Open collaboration has the ability to yield innovative solutions to complex problems. However, not all platforms or collaborative relationships in digital ecosystems lead to consistent outcomes (Mergel, 2015). Using established computational social science methods and topic modeling, an exploratory analysis of the emerging topics related to "open" climate science in repositories on GitHub will be conducted. New research on open collaboration in GitHub may lead to insights on mechanisms supporting effective collaboration for urgent global challenges. The dataset will be pulled using the GitHub API. A primary focus will be determining the specific repositories to use for the topic modeling (Russell et al., 2018). Pull requests and forks of repositories, along with adding issues and other contributions will be included in the topic modeling analysis (Brisson et al., 2020). These data on developer-follower collaborations can be obtained using various GET requests (Bana and Arora, 2018). The anticipated contributions include further understanding the structure of open collaboration in digital ecosystems, and developing an understanding of how data science collaboration and/or theory-guided data science emerges in open collaboration environments (Karpatne et al., 2017; Zhang et al., 2019). As new methods and tools emerge, exploratory studies can help develop an understanding of these themes and relationships in order to develop more detailed studies.

References

Bana, R. & Arora, A. (2018). Influence indexing of developers, repositories, technologies and programming languages on social coding community GitHub. *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 1-6. https://www.doi.org/10.1109/IC3.2018.8530644

Brisson, S., Noei, E., & Lyons, K. (2020). We are family: Analyzing communication in GitHub software repositories and their forks. *2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 59-69, doi: 10.1109/SANER48275.2020.9054834.

Karpatne, A., et al. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *In IEEE Transactions on Knowledge and Data Engineering, 29*(10), 2318- 2331. https://doi.org/10.1109/TKDE.2017.2720168

Mergel, I. (2015). Open collaboration in the public sector: The case of social coding on GitHub. *Government Information Quarterly, 32*(4), 464-472.

Russell, P.H., Johnson, R.L., Ananthan, S., Harnke, B., & Carlson, N.E. (2018) A large-scale analysis of bioinformatics code on GitHub. *PLoS ONE, 13*(10): e0205898. https://doi.org/10.1371/journal.pone.0205898

Zhang, A. X., Muller, M. & Wang, D. (2019). How do data science workers collaborate? Roles, workflows, and tools. *Proceedings of the ACM Human Computer Interaction, 4*. https://doi.org/10.1145/339286

Presentation at TREO Talks in conjunction with the 42nd International Conference on Information Systems, ICIS 2021
TREO Talks are not peer-reviewed and not a formal part of the ICIS 2021 Proceedings
All TREO Talks are available in the TREO Talks section of the AIS e-Library