The Journal of Open Source Software

SWORDS: A framework for the Scan and revieW of Open Research Data and Software

Keven Quach [1*¶], **Jonathan de Bruin** [2*], **and Anna-Lena Lamprecht** [3*]

1 Independent Researcher, Germany **2** Utrecht University, Netherlands **3** University of Potsdam, Germany ¶ Corresponding author * These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

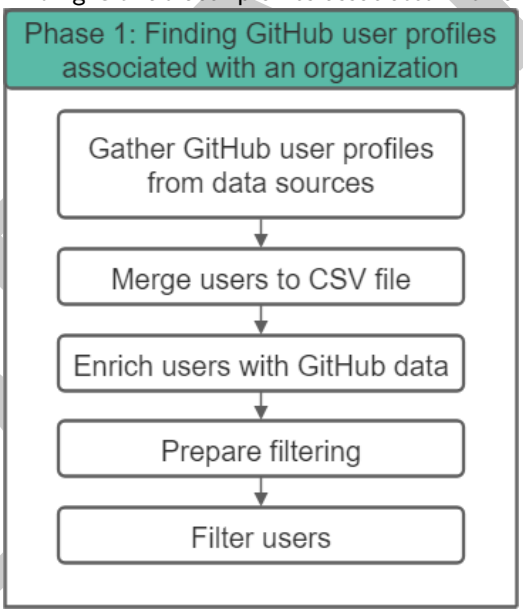
- @openjournals

Submitted: 01 January 1970
Published: unpublished

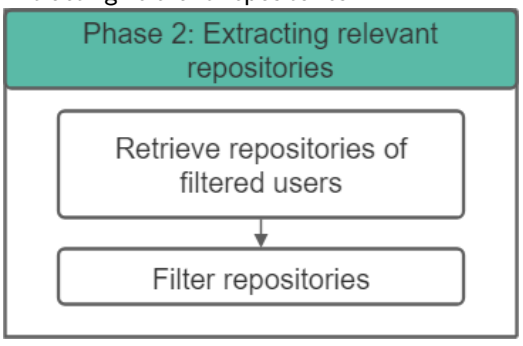
Summary

SWORDS (Scan and revieW of Open Research Data and Software) is a framework designed to provide insights into an organization's, as well as their members open-source activities, through a structured approach. The framework focuses on organizations within the research domain by taking academic publishing principles into account. A big challenge for such a framework lies in the decentralization of open-source activities. It is divided into three core stages that can be executed independently:

1. Finding GitHub user profiles associated with an organization.

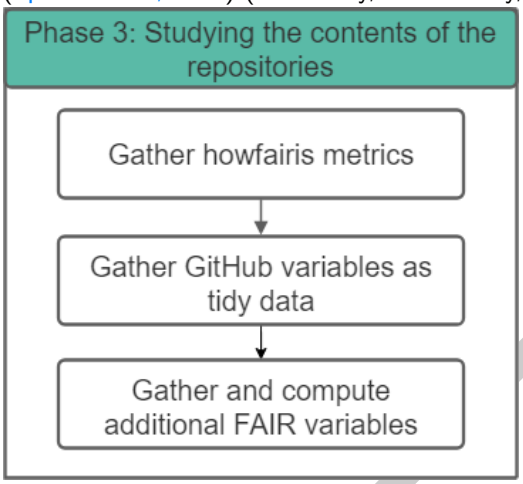


2. Extracting relevant repositories.



The Journal of Open Source Software

3. Studying the contents of the repositories. Content evaluation includes aspects of quality assessment, documentation availability, and FAIRness (Wilkinson et al., 2016) scores (Spaaks et al., 2022) (Findability, Accessibility, Interoperability, and Reusability).



To illustrate, an organization that already has collected the GitHub user profiles does not need to execute phase 1. An organization that is only interested in collecting the GitHub user profiles does not need to execute the following phases. An organization that has already collected relevant repositories does not need to execute phases 1 and 2.

In the past, research organizations had insufficient tooling for the analysis of research output like software and data. Over the years, many initiatives were introduced and especially the introduction of the FAIR principles contributed to an improvement of available tools (Barker et al., 2022). This enabled us to have a better understanding of the publication principles for transparent and reproducible publication of data and software that is also quantifiable through monitoring and evaluation. These insights can be useful as described in the statement of need.

Written in Python, SWORDS provides a template for easy implementation within any organization and focuses on GitHub, which is the go-to reference for mining open-source repositories (Cosentino et al., 2017). It is designed to be extensible and flexible, which allows to evaluate repositories on custom-defined metrics and collect users according to different strategies. The framework was applied to Utrecht University as part of a research project (Quach, 2022). The results of the aforementioned research project were presented at a conference (Quach et al., 2023). Ongoing research projects are currently being conducted at the University of Potsdam. There are also related open source program office (OSPO) tools, which do not fit the academic use-case yet.

Statement of need

Open Science, promoting transparency in academic publications, data, software, and other types of output, is crucial for enhancing scientific and societal impact in today's research climate. The application of Open Science principles to research data and software is vital for ensuring scientific integrity and reproducibility, which can sometimes be lackluster (Allison et al., 2016). However, substantial challenges persist in tracking, managing, and understanding open-source research software due to the scattered and fragmented nature of these activities across multiple platforms (Lamprecht et al., 2020).

The SWORDS framework addresses this need by providing a systematic approach to collating, analyzing, and understanding an organization's open-source research software. The insights gained from implementing SWORDS can help organizations connect initiatives, improve quality, reward and recognize contributions, and foster a collaborative and open research environment. Thus, SWORDS presents an invaluable tool for any research organization aiming to improve

and better understand its open-source activities and drive forward Open Science.

Acknowledgements

We acknowledge contributions from Christopher Slewe during the genesis of this project. We also acknowledge the funding programs: The Open Science Programme (OSP) and FAIR Research IT.

References

Allison, D. B., Brown, A. W., George, B. J., & Kaiser, K. A. (2016). Reproducibility: A tragedy of errors. *Nature*, *530*(7588), 27–29. https://doi.org/10.1038/530027a

Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., & Honeyman, T. (2022). Introducing the FAIR principles for research software. *Scientific Data*, *9*(1), 622. https://doi.org/10.1038/s41597-022-01710-x

Cosentino, V., Cánovas Izquierdo, J. L., & Cabot, J. (2017). A Systematic Mapping Study of Software Development With GitHub. *IEEE Access*, *5*, 7173–7192. https://doi.org/10.1109/ACCESS.2017.2682323

Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., Sandt, S. van de, Ison, J., Martinez, P. A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J. Ll., Chue Hong, N., Goble, C., & Capella-Gutierrez, S. (2020). Towards FAIR principles for research software. *Data Science*, *3*(1), 37–59. https://doi.org/10.3233/DS-190026

Quach, K. (2022). *Mapping Research Software Landscapes through Exploratory Studies of GitHub Data* [Master's thesis]. https://studenttheses.uu.nl/handle/20.500.12932/43162

Quach, K., Lamprecht, A.-L., & De Bruin, J. (2023). *Mapping Research Software Landscapes through Exploratory Studies of GitHub Data*. Zenodo. https://doi.org/10.5281/zenodo.8150215

Spaaks, J. H., Verhoeven, S., Tjong Kim Sang, E., Diblen, F., Martinez-Ortiz, C., Etuk, E., Kuzak, M., Werkhoven, B. van, Soares Siqueira, A., Saladi, S., & Holding, A. (2022). *howfairis* (Version 0.14.2). https://github.com/fair-software/howfairis

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18