

# Promoting research through quality-controlled data and code publications: Collaboration between research and infrastructure

Sarbani Roy<sup>1,2,\*</sup>, Jan Range<sup>1</sup>, Sibylle Hermann<sup>1,2</sup>, Fang Fang Wang<sup>1</sup>, Jenny Kopsch-Xhema<sup>1</sup>

<sup>1</sup> SimTech, University of Stuttgart, Germany

<sup>2</sup> FoKUS, University of Stuttgart, Germany

\* `sarbani.roy@simtech.uni-stuttgart.de`

The digital transformation in science has led to an increased focus on the publication of research data and software, as well as their quality assurance. The traditional method of measuring scientific performance by the number and citation of scientific articles is not sufficient, and there is a growing recognition for appropriately acknowledging the work for well-reusable datasets or research software [1, 2, 3, 4]. Initiatives such as Make Data Count [5] are working to establish an infrastructure for citations to data publications. This trend indicates that high-quality, easily citable, and reusable data and software publications will play an increasingly important role in the evaluation and funding of research. The Research Data Management (RDM) team of the Cluster of Excellence SimTech collaborates with the Research Data Competence Center FoKUS to ensure the quality of research data and software [6].

To achieve this goal of high quality, research data and software are published in a citable form on the institutional research data repository DaRUS. The publication process undergoes a two-stage quality check, first by the SimTech RDM team and then formally optimized by the FoKUS team, to ensure quality review and make it linkable in terms of a persistent PID graph [7] to uniquely identify a publication, dataset, or author. Persistent identifiers can be linked unambiguously using their metadata, regardless of whether they are of the same type (such as journal articles citing other journal articles) or different types (such as a researcher and the datasets they created). The team is currently working on automating this review process as much as possible and improving it through content-related components of the quality check. Additionally, a showcase is being developed to demonstrate the suggested way to prepare a dataset for research data and software.

Archiving research software requires additional measures, such as ensuring the reproducibility of the data generated by the software and the reusability of the software itself. Containerization technology is commonly used to ensure the reproducibility and easy-sharing of software. To facilitate the archivability of research software, a container registry is being developed to host the images of the containers, along with a metadata block to regenerate the run-command when the container was built. We also ensure that the archived research software follows the best practices in software development [8]. To provide further support to researchers, we have developed a tool for creating, uploading, downloading, and editing existing entries of datasets, and we provide consultation services for RDM queries.

SimTech is leading the efforts in this field, and the processes and tools developed will be rolled out across the university. In the future, individual quality criteria will be supported, and badges will be used to make them visible. The integration of Make Data Count in DaRUS will prepare metrics based on the citation of published datasets, which will provide technical capabilities for metrics around research data (and software) on the infrastructure side and give these metrics a higher priority from a SimTech perspective.

## References

- [1] UNESCO. 2021. UNESCO Recommendation on Open Science. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>
- [2] Wissenschaftsrat. 2020. Zum Wandel in den Wissenschaften durch datenintensive Forschung, Wissenschaftsrat. [https://www.wissenschaftsrat.de/download/2020/8667-20.pdf?\\_\\_blob=publicationFile&v=5](https://www.wissenschaftsrat.de/download/2020/8667-20.pdf?__blob=publicationFile&v=5)
- [3] DFG. 2022. Open Science als Teil der Wissenschaftskultur. Positionierung der Deutschen Forschungsgemeinschaft". *Zenodo*. doi: 10.5281/zenodo.7193838.
- [4] DFG— AG Publikationswesen. 2022. Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung. *Zenodo*. doi: 10.5281/zenodo.6538163.
- [5] Make Data Count. <https://makedatacount.org/>
- [6] C. Brecher et al. 2022. Commitment zu aktivem daten- und -softwaremanagement in großen forschungsverbänden: Commitment to active data and software management in large research alliances. *Bausteine Forschungsdatenmanagement*. pp. 121–123. doi: 10.17192/bfdm.2022.1.8412.
- [7] M. Fenner, and A. Aryani. 2019. Introducing the PID Graph. doi:10.5438/JWVF-8A66
- [8] R. C Jiménez et al. 2017. Four simple recommendations to encourage best practices in research software. *F1000Res* doi: 10.12688/f1000research.11407.1