

CiTAR - Citing & Archiving Research

Felix Bartusch, Jens Kruger**

High-Performance and Cloud Computing Group
Zentrum für Datenverarbeitung

University of Tübingen, Germany

*jens.krueger@uni-tuebingen.de

Klaus Rechert, Oleg
Zharkov
University of Freiburg

Kyryll Udod
University of Ulm

ABSTRACT

The importance of software and web services is changing. So it sufficed to state the software used during scientific data processing, when publishing results in journals. This also applies to web services, which can be methods used in scientific work as well as results of this work. Awareness for the evanescence of tools and services, which hinders the reproducibility of published work, increases.

Similar to data repositories, which archive datasets and make them citable, a service for software artifacts would enhance sustainability of science. So we present the CiTAR (Citing & Archiving Research) service in this paper, which enables researchers to preserve computational environments and make them citable. In contrast to pure data repositories, CiTAR guarantees the executability of archived environments by providing generic runtimes.

Keywords—Virtual Machines, Containerization, Long-term Archiving, Preservation, Reproducibility

Informationsverarbeitung und Kommunikation, vol. 35, no. 4, pp. 259–267, 2013.

- [14] K. Rechert, T. Liebetaut, D. Wehrle, and E. Cochrane, “Preserving Containers – Requirements and a Todo-List,” in *Digital Libraries: Knowledge, Information, and Data in an Open Access Society*, A. Morishima, A. Rauber, and C. L. Liew, Eds. Cham: Springer International Publishing, 2016, pp. 225–230.
- [15] S. Sun, L. Lannom, and B. Boesch, “Handle system overview,” Tech. Rep., 2003.
- [16] I. Podtergera, S. Mocken, and J. Besters-Dilger, *SLaVaCompCOMPutergestützte Untersuchung von Variabilität im KirchenSLAvischen: Forschungsergebnisse*. Albert-Ludwigs-Universität Freiburg, 2016.
- [17] J. Pfeuffer, T. Sachsenberg, O. Alka, and M. Walzer, “OpenMS A platform for reproducible analysis of mass spectrometry data,” *Journal of Biotechnology journal*, vol. 261, no. May, pp. 142–148, 2017.
- [18] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kotter, T. Meinl, P. Ohl, K. Thiel, and B. Wiswedel, “KNIME - the Konstanz information miner,” *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, p. 26, 2009.

REFERENCES

- [1] D. J. Lee and B. Stvilia, “Practices of research data curation in institutional repositories: A qualitative view from repository staff,” *PLoS ONE*, vol. 12, no. 3, pp. 1–44, 2017.
- [2] “Software with impact,” *Nature Methods*, vol. 11, no. 3, pp. 211–211, 2014.
- [3] K. Hinsén, “Platforms for publishing and archiving computer-aided research,” *F1000Research*, vol. 3, no. 289, pp. 1–18, 2014.
- [4] D. Forschungsgemeinschaft, *Sicherung guter wissenschaftlicher Praxis (Safeguarding Good Scientific Practice)*, 2013.
- [5] D. Garijo, S. Kinnings, L. Xie, L. Xie, Y. Zhang, P. E. Bourne, and Y. Gil, “Quantifying reproducibility in computational biology: The case of the tuberculosis drugome,” *PLoS ONE*, vol. 8, no. 11, pp. 1–11, 2013.
- [6] R. D. Peng, “Reproducible research in computational science,” *Science*, vol. 334, no. 6060, pp. 1226–1227, 2011.
- [7] D. S. Katz, K. E. Niemeyer, and A. M. Smith, “Strategies for Biomedical Software Management, Sharing, and Citation,” *PeerJ Preprints*, 2016. [Online]. Available: <https://peerj.com/preprints/2640.pdf>
- [8] D. S. Katz, “Software Citations and the ACAT Community,” in *Journal of Physics: Conference Series*, vol. 1085, no. 2, 2018.
- [9] M. e. Jackson, “Checklist for a Software Management Plan,” 2018. [Online]. Available: <https://zenodo.org/record/2159713/files/XDSdeiDjJEY>
- [10] V. V. Sochat, C. J. Prybol, and G. M. Kurtzer, “Enhancing reproducibility in scientific computing: Metrics and registry for Singularity containers,” *PLoS ONE*, vol. 12, no. 11, pp. 1–24, 2017.
- [11] P. Ewels, A. Peltzer, D. Moreno, rfenouil, M. Garcia, S. Panneerselvam, marchoepfner, jun wan, S. F., aanil, S. Haglund, A. Jemt, P. D. Tommaso, L. Veeravalli, J. Alneberg, C. Davenport, R. Suhecki, M. N. Adulyanukosol, Francesco, and C. Wang, “nf-core/rnaseq: nf-core/rnaseq version 1.0,” Aug. 2018. [Online]. Available: <https://doi.org/10.5281/zenodo.1400711>
- [12] F. Bartusch, M. Hanussek, and J. Kruger, “Containerization of Galaxy Workflows increases Reproducibility,” *Proceedings of the bwHPC Symposium*, pp. 16–19, 2017. [Online]. Available: <https://publikationen.uni-tuebingen.de/xmlui/handle/10900/83810>
- [13] K. Rechert, I. Valizada, D. von Suchodoletz, and J. Latocha, “bwFLA A Functional Approach to Digital Preservation,” *PIK - Praxis der*