

CiTAR - Citable Scientific Software and Software Methods

Klaus Rechert, Rafael Gieschke, Susanne Mocken, Oleg Stobbe, Oleg Zharkov (University Freiburg), Felix Bartusch (University Tübingen), Kyrill Udod (University Ulm)

Today's research projects do not only rely on computer assisted methodology and digital resources, but increasingly devote significant resources into creating new digital resources and methods.

A pillar of the scientific work is the reproducibility of published results. For computational methods it is often difficult or even impossible to recreate the exact same software stack of the published results [1]. More and more journals require to provide source code and other computational resources as additional material to publish a paper [2].

CiTAR¹ - Citing and Archiving Research, a 3 year Baden-Württemberg state project, is developing infrastructure to support computer assisted research. A major outcome of the project are means to publish, cite and provide long-term access to virtual research environments e.g. form of virtual machines and scientific processes in form of containers.

Virtual research environments or similar public facing services are typical outcomes of research projects. Their long-term availability and accessibility is however an open question. The costs for maintaining a server and/or software after the end of a project are usually not covered, especially if these projects are only of interest for a small and specialized research community. Leaving an unmaintained, outdated machine connected to the internet poses a latent and increasing security risk.

As part of this process, a CiTAR publisher is able to upload an image (e.g. a VM) which is then in a fully automated process analyzed and technically normalized, to ensure future compatibility with emulation or virtualization infrastructure and thus ensure long-term preservation options [3].

Published, archived and in particular citable research environment imply long-term access. Hence the second challenge of the CiTAR project was efficient and secure access to the archived machine. As this machine has been archived to remain in its original state, a (permanent) internet connection could be harmful. Furthermore, most of these machines target small groups and won't be accessed that often. We addressed this problem set by on-demand, cloud-based provisioning, i.e. a visitor of the respective landing page initiates the startup of the instance. In order to provide (secure) network access [4].

Furthermore, container technology has been quickly adopted as a tool to encapsulate and share complex software setups, e.g. in the domain of computational science. With growing significance of this class of complex digital objects their longevity is also of growing importance. We have analysed a container's long-term preservation risks [5] and have developed a normalization process to unify container's to single homogenous class of digital objects. Based on this, we have implemented an emulation-based preservation strategy -- converting them into a generic archival representation and providing a generic runtime

¹ Project website: <http://citar.eaas.uni-freiburg.de/>

environment -- as well as necessary workflows to maintain access to software-based research methods.

[1] Garijo, D., Kinnings, S., Xie, L., Xie, L., Zhang, Y., Bourne, P. E., & Gil, Y. (2013). Quantifying reproducibility in computational biology: The case of the tuberculosis drugome. *PLoS ONE*, 8(11), 1–11. <https://doi.org/10.1371/journal.pone.0080278>

[2] Software with impact. (2014). Nature Publishing Group, 11. <https://doi.org/10.1038/nmeth.2880>

[3] Rechert, K., Falcao, P., & Ensom, T. (2016). Towards a Risk Model for Emulation-based Preservation Strategies: A Case Study from the Software-based Art Domain. In *Proceedings of the 13th International Conference on Digital Preservation* (pp. 139-148).

[4]<http://openpreservation.org/blog/2018/12/12/preserving-virtual-research-environments-introducing-citar-part-1/>

[5] Rechert, K, Liebetraut, T., Wehrle, D., and Euan C.. "Preserving Containers—Requirements and a Todo-List." In *International Conference on Asian Digital Libraries*, pp. 225-230. Springer, Cham, 2016.