

## Community Expectations for Research Artifacts and Evaluation Processes

Ben Hermann<sup>1</sup>, Stefan Winter<sup>2</sup>, Janet Siegmund<sup>3</sup>

**Abstract:** Artifact evaluation has been introduced into the software engineering and programming languages research community with a pilot at ESEC/FSE 2011 and has since then enjoyed a healthy adoption throughout the conference landscape. We conducted a survey including all members of artifact evaluation committees of major conferences in the software engineering and programming language field from 2011 to 2019 and compared the answers to expectations set by calls for artifacts and reviewing guidelines. While we find that some expectations exceed the ones expressed in calls and reviewing guidelines, there is no consensus on a quality threshold for artifacts in general. We observe very specific quality expectations for specific artifact types for review and later usage, but also a lack of their communication in calls. We also find problematic inconsistencies in the terminology used to express artifact evaluation's most important purpose. We derive several actionable suggestions which can help to mature artifact evaluation in the inspected community and also to aid its introduction into other communities in computer science.

**Keywords:** Research Artifacts; Artifact Evaluation; Replicability; Reproducibility; Study

### 1 Summary

In this paper, we present a study on the expectations of the community toward research artifacts and their evaluation processes which was originally presented at ESEC/FSE 2020 and has received the ACM SIGSOFT Distinguished Paper Award [HWS20]. Since the publication of the original study it had impact on numerous artifact evaluation tracks and inspired a quantitative study on artifact quality and visibility published at ESEC/FSE 2022 [Wi22].

In 2016, a replicability crisis became public, when more than 1500 researchers revealed having trouble replicating previous research results. This replicability crisis also reached the software engineering community, as it has embraced the importance of replication for knowledge building. To improve the situation of missing or unusable research artifacts, artifact evaluation has become a regular process for scientific conferences in computer

---

<sup>1</sup> Technische Universität Dortmund, Fakultät für Informatik, Otto-Hahn-Straße 14, 44227 Dortmund, Deutschland  
ben.hermann@cs.tu-dortmund.de

<sup>2</sup> Ludwig-Maximilians-Universität München, Fakultät für Informatik, Oettingenstraße 67, 80538 München, Deutschland sw@stefan-winter.net

<sup>3</sup> Technische Universität Chemnitz, Fakultät für Informatik, Straße der Nationen 62, 09111 Chemnitz, Detuschland  
janet.siegmund@informatik.tu-chemnitz.de

science. Since the first piloting of the process at ESEC/FSE 2011, many other conferences have included artifact evaluations as an additional step.

The overarching goal of our work is to enable an assessment of the efficacy of artifact evaluations and to identify possible improvements for these processes. As a first step towards a systematic assessment of artifact evaluation processes, the objective of our work is to assess their current perception in the AE-pioneering communities. We have conducted a survey among researchers who have served on artifact evaluation committees (AECs). To this end, we have contacted all members of AECs, including the respective chairs, for all artifact evaluations conducted at software engineering and programming language conferences between 2011 and 2019.

We found that the perceived purpose of artifact evaluation is to foster replicability and reusability at the same time. While we could observe several quality criteria to be expected from artifacts, we found no clear consensus on them. Moreover, the expressed expectations were largely not represented in the calls for artifacts. This makes it hard to define a quality standard. The results of our study show that the lack of such quality standards leaves reviewers without guidance how to decide on artifact acceptance or rejection. Moreover, it creates an ambiguity for readers how to interpret the badges awarded to papers after AE.

In summary, we make the following contributions: (1) We provide an overview of the current perception and practice of artifact evaluation and the expectations toward artifacts and the process. (2) Based on community inputs, we present suggestions for future development and improvement of artifact evaluations. (3) We published a research artifact for replication, further analysis, and extension by the community.

## 2 Data Availability

The original publication is accessible under the DOI [10.1145/3368089.3409767](https://doi.org/10.1145/3368089.3409767) [HWS20]. Our artifact is available on Github (<https://github.com/bhermann/artifact-survey>) and Zenodo (DOI: [10.5281/zenodo.3951724](https://doi.org/10.5281/zenodo.3951724)).

## Bibliography

- [HWS20] Hermann, Ben; Winter, Stefan; Siegmund, Janet: Community Expectations for Research Artifacts and Evaluation Processes. In: Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ESEC/FSE 2020, Association for Computing Machinery, New York, NY, USA, p. 469–480, 2020.
- [Wi22] Winter, Stefan; Timperley, Christopher S.; Hermann, Ben; Cito, Jürgen; Bell, Jonathan; Hilton, Michael; Beyer, Dirk: A Retrospective Study of One Decade of Artifact Evaluations. In: Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ESEC/FSE 2022, Association for Computing Machinery, New York, NY, USA, p. 145–156, 2022.