Replication Package of "Evaluation Methods and Replicability of Software Architecture Research Objects"

Marco Konersmann* Angelika Kaplan† Thomas Kühn† Robert Heinrich† Anne Koziolek† Ralf Reussner†
Jan Jürjens*‡ Mahmood al-Doori* Nicolas Boltz† Marco Ehl* Dominik Fuchß† Katharina Großer* Sebastian Hahner†
Jan Keim† Matthias Lohr* Timur Sağlam† Sophie Schulz† Jan-Philipp Töberg†§
*University of Koblenz-Landau, Germany, {lastname,mahmoodaldoori,mehl,matthiaslohr}@uni-koblenz.de
†Karlsruhe Institute of Technology, Germany, {firstname.lastname}@kit.edu, §uexdy@student.kit.edu
‡Fraunhofer Institute for Software and Systems Engineering, ISST, Germany

Abstract—In "Evaluation Methods and Replicability of Software Architecture Research" [1], we present a systematic literature review to assess the state-of-practice of evaluating software architecture research objects and providing replication artifacts in 153 full technical conference papers published at the International and European Conference on Software Architecture (ICSA respectively ECSA) from 2017 to 2021.

Context: Software architecture (SA) as research area experienced an increase in empirical research. Empirical research builds a sound foundation for the validity and comparability of the research. A current overview on the evaluation and replicability of SA research objects could help to discuss our empirical standards as a community. However, no such current overview exists.

Objective: We aim at assessing the current state of practice of evaluating SA research objects and replication artifact provision in full technical ICSA and ECSA conference papers from 2017 to 2021.

Method: We first create a categorization of papers regarding their evaluation and provision of replication artifacts. In a systematic literature review (SLR) with 153 papers we then investigate how SA research objects are evaluated and how artifacts are made available.

Results: We found that technical experiments (28%) and case studies (29%) are the most frequently used evaluation methods over all research objects. Functional suitability (46% of evaluated properties) and performance (29%) are the most evaluated properties. 17 papers (11%) provide replication packages and 97 papers (63%) explicitly state threats to validity. 17% of papers reference guidelines for evaluations and 14% of papers reference guidelines for threats to validity.

Conclusions: Our results indicate that the generalizability and repeatability of evaluations could be improved to enhance the maturity of the field; although, there are valid reasons for contributions to not publish their data. We derive from our findings a set of four proposals for improving the state of practice in evaluating software architecture research objects. Researchers can use our results to find recommendations on relevant properties to evaluate and evaluation methods to use

and to identify reusable evaluation artifacts to compare their novel ideas with other research. Reviewers can use our results to compare the evaluation and replicability of submissions with the state of the practice.

Replication Package: This package contains the data produced and scripts used during the literature review:

- **Investigated Papers.bib**: A BibTeX file with all papers investigated in the paper.
- **Raw-Data-Table Content-Data.html and Raw-Data-Table Meta-Data.html**: Contains tables with the raw data as extracted during the systematic literature review.
- **Collection of Data Visualizations.pdf**: Shows multiple visualizations of the raw data for analysis. This is a copy of the `summary.pdf` as described below.
- **Data and Visualization** contains:
  - the data as CSV files,
  - scripts for creating visualizations (including AWK, Ruby, LaTeX, and Shell scripts).
  - `README.md` – describing the contained scripts.
  - `summary.pdf` – collection of diagrams (as built by `make-all.sh`)
  - `paper-figures.pdf` – collection of those diagrams used in the accompanying paper (as built by `make-paper-figures.sh` or `make-all.sh`)
- **Wiki**: A copy of the wiki used during data extraction with descriptions of (1) all data items, (2) the process and (3) the employed taxonomy.

The package is publicly available online:

- on Zenodo: https://doi.org/10.5281/zenodo.6044059 (License CC-BY-A-4.0) and
- on GitLab for convenient collaboration: https://gitlab.com/SoftwareArchitectureResearch/StateOfPractice

REFERENCES

[1] M. Konersmann, A. Kaplan, T. Kühn, R. Heinrich, A. Koziolek, R. Reussner, J. Jürjens, M. al Doori, N. Bolz, M. Ehl, D. Fuchß, K. Großer, S. Hahner, J. Keim, M. Lohr, T. Sağlam, S. Schulz, and J.-P. Töberg, "Evaluation Methods and Replicability of Software Architecture Research Objects," in *ICSA 2022*.