Artifact Evaluation Practices in SE/PL vs. Systems

Stefan Winter
sw@stefan-winter.net

LMU Munich

Sep 29 2023

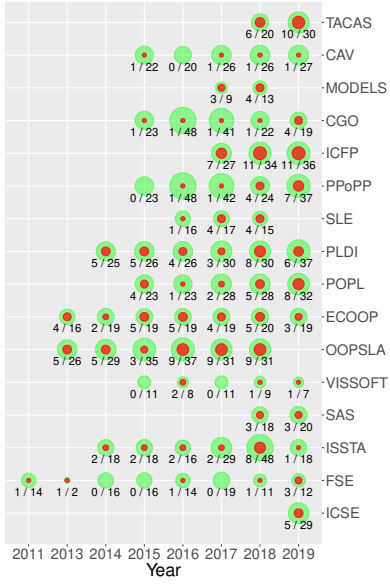
Presentation Context

Increasing Adoption of AE in Systems Conferences:

- SOSP since 2019
- ASPLOS since 2020
- OSDI since 2020
- EuroSys since 2021
- SC since 2021
- USENIX ATC since 2022

Objectives:

1. Share insights from studying SE/PL AEs (for AE organizers and artifact authors)
2. Call for collaborations: Reproduce research results



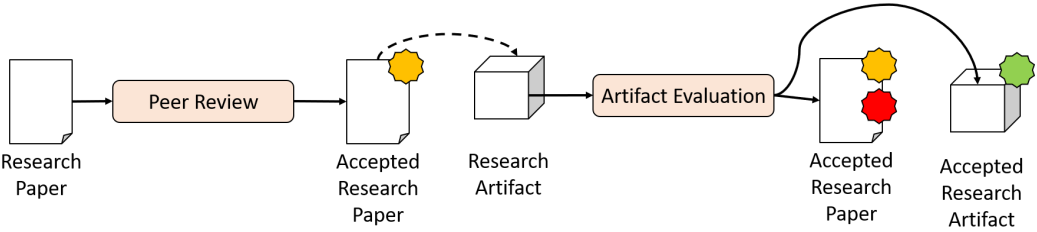
Artifact Definitions

ACM: "a digital object that was either created by the authors to be used as part of the study or generated by the experiment itself"
https://www.acm.org/publications/policies/artifact-review-and-badging-current
Also applies for SOSP, EuroSys

USENIX: No definition, but list of examples: "software, hardware, evaluation data and documentation, raw survey results, mechanized proofs, models, test suites, benchmarks, and so on"
Usually excluded: hardware, paper proofs

SC: No definition, but list of examples: software, datasets, environment configuration, mechanized proofs, benchmarks, test suites with scripts, etc.

Artifact Evaluations



Does the artifact work?

Is the artifact permanently available?

Can the results be confirmed?

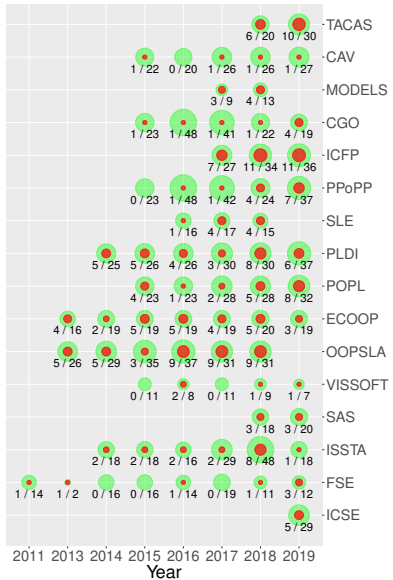
What Makes a *Good* Artifact? [FSE'20]

26 survey questions covering...

- Purpose and value of artifact evaluations
- Criteria for accepting/rejecting artifacts
- Expectations on (re-)used artifacts

Few *general* quality attributes:

- Documentation
- Availability
- Reproducibility



Do Artifact Evaluations Yield Better Artifacts? [FSE'22]

RQ1: Are articles with artifacts that have passed AE more visible?

RQ2: Are successfully evaluated artifacts more available?

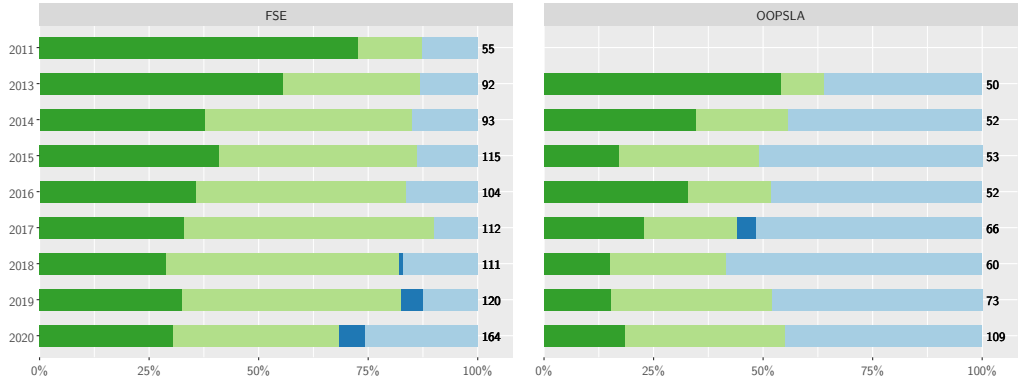
RQ3:Is artifact development/maintenance continued more often for successfully evaluated artifacts?

RQ4: Are successfully evaluated artifacts more often reused?

RQ5: Are successfully evaluated artifacts more thoroughly documented?

Subjects and Initial Classification

Subjects: 3,650 research articles from 64 proceedings across 12 SE/PL conferences



RQ2: AE Effects on Artifact Availability

AE Evaluated	Available Badge Status	Total Papers	Has Artifact Reference	Is Accessible
AE	Av. Badge	683	676 (99.0%)	675 (99.9%)
	No Av. Badge	602	473 (78.6%)	431 (91.1%)
NonAE	Av. Badge	71	67 (94.4%)	65 (97.0%)
	No Av. Badge	2294	1148 (50.0%)	1032 (89.9%)

"Available" badges are positively linked with both reference accessibility and artifact evaluations

RQ5: AE Effects on Documentation

- Random sample of 100 AE and 100 NonAE artifacts
- Search documentation files by regex

Search Term	Matched Artifacts		Word Count	
	AE	NonAE	AE	NonAE
^read.*me	84	86	1,389	645
^install	6	1	324	593
^doc/	1	8	2,431	13,901
No match	13	12	–	–
^copyright	0	1	0	268
^license	50	46	850	1,220

Documentation practices differ. Licenses and copyright information are often missing.

Summary of Findings so Far

- Availability: Use Permanent Archives
  - Anything that issues a DOI works (Zenodo, Figshare, Dryad, ...)
  - USENIX currently accepts GitHub/Gitlab
  - Propagates back to ACM conferences, but opposes ACM definitions

Year	Available	Unavailable
2019	21	1
2020	46	5
2021	280	3
2022	185	7

- Documentation is lacking standards – ACM SIGSOFT is now converging to:
  - `README`: What is it?
  - `REQUIREMENTS`: What is needed to operate it?
  - `STATUS`: Which badges are claimed why?
  - `LICENSE`: Terms for (re-)use
  - `INSTALL`: Setup instructions

Reproducibility (ongoing)

Reproducing research results in seminars

1. Read/discuss background literature (4 weeks)
2. *Reproduce research results (6 weeks)*
3. Consolidate & present results (2 weeks)

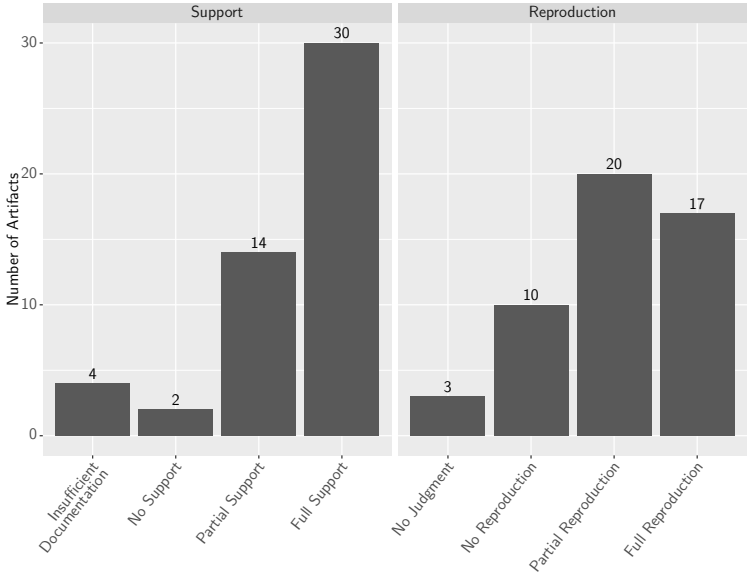
Growing number of participating universities

- LMU Munich
- TU Dortmund
- TU Wien
- HU Berlin
- Your institution here?

Preliminary Results (LMU, WiSe 2021/22): Initial Filters

Status	Count
Broken SW dependency	1
Unavailable	3
Timeout	3
Unmet HW dependencies	5
Processed	50
Total	62

Preliminary Results (LMU, WiSe 2021/22): Reproducibility



Call for Collaborations

- System artifacts LMU from WiSe 2024
- I am happy to share seminar resources & experiences