

# Artifact Evaluation for Publications

Edited by

Bruce R. Childers<sup>1</sup>, Grigori Fursin<sup>2</sup>, Shriram Krishnamurthi<sup>3</sup>, and  
Andreas Zeller<sup>4</sup>

- 1 University of Pittsburgh, US, [childers@cs.pitt.edu](mailto:childers@cs.pitt.edu)
- 2 cTuning - Cachan, FR, [grigori.fursin@ctuning.org](mailto:grigori.fursin@ctuning.org)
- 3 Brown University - Providence, US, [sk@cs.brown.edu](mailto:sk@cs.brown.edu)
- 4 Universität des Saarlandes, DE, [zeller@cs.uni-saarland.de](mailto:zeller@cs.uni-saarland.de)

---

## Abstract

---

This report documents the program and the outcomes of Dagstuhl Perspectives Workshop 15452 “Artifact Evaluation for Publications”. This Perspectives Workshop conveyed several stakeholders in artifact evaluation from different communities to assess how artifact evaluation is working and make recommendations to the computer systems research community about several issues with the process.

**Seminar** November 1–4, 2015 – <http://www.dagstuhl.de/15452>

**1998 ACM Subject Classification** H.3.7 Digital Libraries, H.3.5 Online Information Services, D.2.9 Management

**Keywords and phrases** Computer systems, artifacts, reproducibility, archive

**Digital Object Identifier** 10.4230/DagRep.5.11.29


## 1 Executive Summary

*Bruce R. Childers*

*Grigori Fursin*

*Shriram Krishnamurthi*

*Andreas Zeller*

**License**  Creative Commons BY 3.0 Unported license  
© Bruce R. Childers, Grigori Fursin, Shriram Krishnamurthi, and Andreas Zeller

Computer systems researchers have developed numerous artifacts that encompass a broad collection of software tools, benchmarks, and data sets. These artifacts are used to prototype innovations, evaluate trade-offs and analyze implications. Unfortunately, methods used in the evaluation of computing system innovation are often at odds with sound science and engineering practice. The ever-increasing pressure to publish more and more results poses an impediment to *accountability*, which is a key component of the scientific and engineering process. Experimental results are not usually disseminated with sufficient metadata (i.e., software extensions, data sets, benchmarks, test cases, scripts, parameters, etc.) to achieve repeatability and/or reproducibility. Without this information, issues surrounding trust, fairness and building on and comparing with previous ideas becomes problematic. Efforts in various computer systems research sub-communities, including programming languages/compilers, computer architecture, and high-performance computing, are underway to address the challenge.

This Dagstuhl Perspectives Workshop (PW) brought together stakeholders of associated CSR sub-communities to determine synergies and to identify the most promising directions



Except where otherwise noted, content of this report is licensed  
under a Creative Commons BY 3.0 Unported license

Artifact Evaluation for Publications, *Dagstuhl Reports*, Vol. 5, Issue 11, pp. 29–35

Editors: Bruce R. Childers, Grigori Fursin, Shriram Krishnamurthi, and Andreas Zeller



DAGSTUHL  
REPORTS

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

and mechanisms to push the broader community toward accountability. The PW assessed current efforts, shared what does and doesn't work, identified additional processes, and determined possible incentives and mechanisms. The outcomes from the workshop, including recommendations to catalyze the community, are separately documented in an associated Dagstuhl Manifesto.

**2 Table of Contents**

**Executive Summary**  
*Bruce R. Childers, Grigori Fursin, Shriram Krishnamurthi, and Andreas Zeller . . .* 29

**Goals . . . . .** 32

**Topics . . . . .** 32

    Assessment of Existing Efforts . . . . . 32

    Pushing the Community Further . . . . . 32

    Building and Sustaining a Community . . . . . 33

    Processes, Mechanisms and Repositories . . . . . 33

**Demos . . . . .** 34

**Participants . . . . .** 35

### 3 Goals

Before the workshop, the organizers identified several goals to engage and drive the event. These goals were:

1. Assess the state of current efforts to achieve accountability, including successes and why these worked, and the impediments being and likely to be faced;
2. Identify strategies and incentives to engage the community and raise expectation for higher experimental quality and accountability;
3. Identify the shape of the most promising approaches for the technical challenges posed by building open-access repositories and associated services;
4. Identify ways to leverage, combine and coordinate existing and new efforts in PL/compiler and software engineering, high-performance computing and computer architecture;
5. Develop recommendations to lead the community to better artifacts and accountable experimental results.

### 4 Topics

To address the goals, the PW was organized around four topics with several questions for participants to consider for each topic. Rather than hold formal talks, the workshop was arranged as a round-table discussion of the topics. The agenda was intentionally informal to allow plenty of time for significant discussion. Participant ideas, hunches, concerns, thoughts, and challenges were every bit as welcome as any particular concrete thing the participants had done. During the discussion, the intent was not to decisively answer the questions – that can only be done with full community involvement over time – but rather to give direction of where answers may be found and how to get there. All of the participants contributed to the discussion. Our informal format was quite successful in creating a lively and productive environment to work through the issues on artifact evaluation.

The specific outcomes (recommendations) from the workshop are described in the Dagstuhl Manifesto for the event. Below, we list the topics and the questions discussed.

#### 4.1 Assessment of Existing Efforts

There are several initial efforts for accountability in CSR sub-communities for PL/compiler and software engineering, computer systems architecture, and HPC. These efforts were discussed to understand how well they are working and what issues are being faced. Some specific questions addressed were:

- What approaches taken by existing efforts have achieved the most traction, and what approaches have faced the most resistance?
- What obstacles have been and will likely be raised?
- What are the similarities and differences among sub-communities?
- What capabilities are currently missing?

#### 4.2 Pushing the Community Further

To reap the benefits of accountability, researchers have to be moved to adopt expectations and procedures for this purpose. Community needs should first be determined, and then

processes and infrastructures can be developed. Incentive is important – individuals must be motivated to participate. Several questions were discussed on how to catalyze the community to do more:

- What is the community’s desired level of trust and leverage from accountability?
- How can the risk of over emphasis on building and evaluating artifacts and experiments be mitigated to avoid imposing too much hassle?
- How can the community be incentivised?
- How do we treat industrial artifacts?
- What is the interaction between evaluation of artifacts and paper acceptance?
- Should journals also participate in these processes?

### 4.3 Building and Sustaining a Community

Widespread community “buy-in” – from funding agencies, to program committees, to journal editors, to individual researchers – is necessary to establish and sustain accountability and associated processes and mechanisms. Accountability must become an inherent expectation for it to be effective and sustainable. Several questions were considered on how a community can be created to support artifact evaluation:

- How can the community be convinced that accountability is beneficial?
- What advocacy strategies will encourage adoption of processes and mechanisms for accountability?
- How can the community be facilitated toward overcoming concerns with privacy/accessibility?

### 4.4 Processes, Mechanisms and Repositories

Accountability relies on access to experimental details, which implies artifacts and associated metadata should be available. To leverage and compare with past innovation in the most effective way requires access to the original prototype implementation. Typically mundane issues with packaging and distribution become vital ones. This topic presented several questions for discussion about what technical capabilities are required:

- Do the AE processes which have been described (and which are working in their areas) work more generally across CS?
  - If so, what needs to be done to have them taken up by the whole community?
  - If not, how do we understand what needs to be changed?
- Is there more that we could do in terms of AE that would improve either trust or usefulness of artifacts?
- How should a repository and associated services be structured? A number of demos and examples were given of existing systems, but an structured analysis of differences and similarities is really required.
- What is a good taxonomy for the repositories and the services?
- How can artifacts, metadata and results be packaged as digital objects for a repository?
- Should journals also participate in these processes? For example, SCP/Elsevier is trying to take a path for artifacts, but the main issue here is “who owns what?” in scientific publications in general, and in artifacts in particular?

## 5 Demos

Several participants gave demos and/or discussions of systems that they have been developing or used for artifact evaluation. These demos included:

- Collective Knowledge, <http://cTuning.org>
- DataMill, <https://uwaterloo.ca/embedded-software-group/datamill>
- Open Curation for Computer Architecture, <http://www.occamportal.org>
- Parallel Workloads Archive, <http://www.cs.huji.ac.il/labs/parallel/workload/>
- Portable Database Files, <http://www.vldb.org/pvldb/vol8/p1972-dittrich.pdf>
- Multi2Sim Heterogeneous System Simulator, <https://www.multi2sim.org/>
- SPADE, <https://github.com/ashish-gehani/spade>
- TIRA – Evaluation as a Service, <http://www.tira.io>

## Acknowledgments

We thank Carlo Ghezzi, Jan Vitek, Matthias Hauswirth, Camil Demetrescu, and the others that have worked with us on artifact evaluation for their input on the process.

## Participants

- Bruce R. Childers  
University of Pittsburgh – Pittsburgh, USA
- Neil Chue Hong  
Software Sustainability Institute – Edinburgh, GB
- Tom Crick  
Cardiff Metropolitan University – Cardiff, GB
- Jack W. Davidson  
University of Virginia – Charlottesville, USA
- Camil Demetrescu  
Sapienza University of Rome – Rome, IT
- Roberto Di Cosmo  
Univ. Paris-Diderot – Paris, FR
- Jens Dittrich  
Universität des Saarlandes – Saarbrücken, DE
- Dror Feitelson  
The Hebrew University of Jerusalem – Jerusalem, IL
- Sebastian Fischmeister  
University of Waterloo – Waterloo, CA
- Grigori Fursin  
cTuning – Cachan, FR
- Ashish Gehani  
SRI – Menlo Park, US
- Matthias Hauswirth  
Univ. of Lugano – Lugano, CH
- Marc Herbstritt  
Schloss Dagstuhl – Wadern, DE
- David R. Kaeli  
Northeastern University – Boston, US
- Shriram Krishnamurthi  
Brown Univ. – Providence, US
- Anton Lokhmotov  
Dividiti Ltd. – Cambridge, GB
- Martin Potthast  
Bauhaus-Universität Weimar – Weimar, DE
- Lutz Prechelt  
FU Berlin, – Berlin, DE
- Petr Tuma  
Charles University – Prague, CZ
- Michael Wagner  
Schloss Dagstuhl – Wadern, DE
- Andreas Zeller  
Universität des Saarlandes – Saarbrücken, DE

