

ICSE'21 Artifact Evaluation Track-Submission and Reviewing Guidelines

Silvia Abrahão and Daniel Mendez

Scope and Objectives

This document outlines, briefly, the submission and reviewing process for the artifact evaluation (AE) track of ICSE 2021. It aims at providing authors and reviewers with pragmatic insights into the process and expected criteria to merit awarding the respective badges.

General Remarks on the AE Track and expected Attitude

In principle, the goal of the track is to promote and celebrate open science. We therefore understand the track as one important means to actively engage with the community in order to support them in making their research artifacts publicly available and in fostering replication of research results. The final result of the artifact evaluation is to reward (only) the authors' work which satisfies the criteria listed below with a set of qualifying badges as a form of recognition. Yet, we see the track and the review phase as a unique chance to actively support the research community in open science, so instead of reviewing the artifacts "blindly" according to the evaluation criteria towards the end of the review phase and submitting a review with a go/no-go decision, we encourage all reviewers to make use of the rebuttal phase in order to actively support the authors in improving their submissions, same as we encourage authors to actively engage with the reviewers and do their best to address their well-intended suggestions as efficiently as possible.

Outline

[Badges Overview and Eligibility](#)

[Submission Process](#)

[Submission Overview](#)

[Types of Research Artifacts](#)

[General documentation](#)

[Review Process](#)

[1. Bidding Phase \(January 16-22\)](#)

[2. Initial Review and Rebuttal Phase \(January 23-February 5\)](#)

[3. In-depth Review Phase \(February 6-21\)](#)

[Summary of Important Dates](#)

[Evaluation Criteria](#)

[Criteria Functional and Reusable badge](#)

[Criteria Available badge](#)





[Criteria Replicated and Reproduced badge](#)

[Further supplementary material to this document \(further reading\)](#)

[Acknowledgements](#)

1.Badges Overview and Eligibility

The artifact evaluation track aims to review, promote, share, and catalog the research artifacts of accepted software engineering papers. Authors of the papers accepted at the Technical, SEIP, NIER, SEET, or SEIS ICSE Tracks can submit an artifact for evaluation as a candidate *Reusable*, *Available*, *Replicated* or *Reproduced* artifact. Authors of any prior SE work (published at ICSE or elsewhere) are also invited to submit their work for evaluation as a candidate for the *Replicated* or *Reproduced* badge. Those badges indicate that the original work has been independently (externally) replicated or reproduced by authors other than those of the original work and will be assigned digitally in retrospect (if supported by the respective publisher). The top two reproduced or replicated artifacts selected by the PC will be awarded the best artifact awards.

Functional	Reusable Open to ICSE 21 submissions only	Available Open to ICSE 21 submissions only	Replicated Open to any submission	Reproduced Open to any submission
No Badge				
Artifacts documented, consistent, complete, exercisable, and include appropriate evidence of verification and validation	Functional + very carefully documented and well-structured to the extent that reuse and repurposing is facilitated. In particular, norms and standards of the research community for artifacts of this type are strictly adhered to.	Reusable + placed on a publicly accessible archival repository. A DOI or link to this repository along with a unique identifier for the object is provided.	Available + main results of the paper have been obtained in a subsequent study by a person or team other than the authors, using, in part, artifacts provided by the author.	Available + the main results of the paper have been independently obtained in a subsequent study by a person or team other than the authors, without the use of author-supplied artifacts.

Please note that we assign either the “Available” badge or the “Reusable” badge, but not both. The “Available” badge is awarded only to artifacts which are also reusable.

Finally, all accepted abstracts documenting the artifacts will be published in the ICSE 2021 proceedings as a further form of recognition.

2. Submission Process

Submission Overview

In principle, authors are expected to submit through [EasyChair](#) their artifact documentation. This documentation distinguishes two basic types of information - **captured in one central research abstract (two pages max)** - depending on the envisioned badge:

1. **Replicated and Reproduced** where the emphasis lies on providing information about how their already published research has been replicated or reproduced as well as links to further material (e.g. the papers and artifacts in question). Note that we encourage submissions for those badges also to nominate other authors (e.g., when authors having reproduced study results want to nominate authors of the original study being replicated/reproduced).
2. **Reusable and Available** where the emphasis lies on providing documentation on the research artifact previously prepared and archived. Here, the authors need to write and submit a documentation explaining how to obtain the artifact package, how to unpack the artifact, how to get started, and how to use the artifacts in more detail. The submission must only describe the technicalities of the artifacts and uses of the artifact that are not already described in the paper.

Note that if the authors are aiming for the badges *Available* and beyond, the artifact needs to be publicly accessible at the time of submission. This means that the EasyChair submission should include the research abstract only providing links to the repositories where the artifact is permanently stored and available. Submitting artifacts themselves through EasyChair without making them publicly accessible (through a repository or an archival service) will not be sufficient for any further badge. In the case of authors applying for the badge *Reusable*, the artifacts do not necessarily have to be publicly accessible for the review process. In this very case, the authors are asked to provide either a private link / password-protected link to a repository or they may submit the artifact directly through EasyChair (in a zip file) and it should become clear which steps are necessary for authors who would like to reuse the artifact.

Details on the research artifacts themselves are provided next.

Types of Research Artifacts

There are two options depending on the nature of the artifacts: Installation Package or Simple Package. If not limited to, while installation packages are typically referred to in context software artifacts or, for instance, scripts, simple packages may be referred to in context of qualitative studies (e.g., interview transcripts or coding schemas).

In both cases, it is expected that the basic set-up of the artifact (including configurations and installations) **do take less than 30 minutes**. Otherwise, the artifact is unlikely to be explicitly endorsed by PC members as they would with other artifacts.

Installation Package. If the artifact consists of a tool or software system, then the authors need to prepare an installation package so that the tool can be installed and run in the

evaluator's environment. That is to say, please make sure to provide enough associated instructions, code, and data such that any Software Engineering person with a reasonable knowledge of scripting, build tools, etc. could install, build, and run the code. If the artifact contains or requires the use of a special tool or any other non-trivial piece of software, the authors must provide a [VirtualBox VM image](#) or a [Docker container image](#) with a working environment containing the artifact and all the necessary tools.

We expect that the artifacts have been vetted on a clean machine before submission.

Simple Package. If the artifact contains documents which can be used with a simple text editor, a PDF viewer, or some other common tool (e.g., a spreadsheet program in its basic configuration) the authors can just save all documents in a single package file (zip or tar.gz). The authors need to make the packaged artifact (installation package or simple package) available so that the Program Committee can access it. We expect that the package is made available through a link to the permanent, public repository, same as it is the case for the installation packages (with the minor exceptions for *Reusable* explained already earlier) and that the archived files are a widely available archive format / platform.

General Documentation

Regardless of the badge, authors must provide documentation explaining how to obtain the artifact package, how to unpack the artifact, how to get started, and how to use the artifacts in more detail. The artifact itself must only describe the technicalities of the artifacts and uses of the artifact that are not already described in the paper; nevertheless, the artifact and its documentation should be self-contained. The submission should contain (and/or link to) the following documents (in plain text or pdf format):

- A README main file describing what the artifact does and where it can be obtained (with hidden links and access password if necessary). Also, there should be a clear description how to repeat/replicate/reproduce the results presented in the paper. Artifacts which focus on data should, in principle, cover aspects relevant to understand the context, data provenance, ethical and legal statements (as long as relevant), and storage requirements. Artifacts which focus on software should, in principle, cover aspects relevant to how to install and use it (and be accompanied by a small example).
- A REQUIREMENTS file for artifacts which focus on software. This file should, in principle, cover aspects of hardware environment requirements (e.g., performance, storage or non-commodity peripherals) and software environments (e.g., Docker, VM, and operating system) but also, if relevant, a requirements.txt with explicit versioning information (e.g. for Python-only environments). Any deviation from standard environments needs to be reasonably justified.
- A STATUS file stating what kind of badge(s) the authors are applying for as well as the reasons why the authors believe that the artifact deserves that badge(s).

- A LICENSE file describing the distribution rights. Note that to score “available” or higher, then that license needs to be some form of open source license. Details also under the respective badges and the ICSE 2021 open science policy.
- An INSTALL file with installation instructions. These instructions should include notes illustrating a very basic usage example or a method to test the installation. This could be, for instance, on what output to expect that confirms that the code is installed and working; and the code is doing something interesting and useful.
- A copy of the accepted paper in pdf format.

3. Review Process

The following section's intended audience is the Program Committee (PC) and, thus, addresses the PC members of the Artifact Evaluation track (and is written accordingly), but it is available to authors as well to facilitate transparency.

The tasks of the reviewers of research artifacts involve three phases:

1. Bidding Phase (January 16-22, 2021)
2. Initial Review and Rebuttal Phase (January 23 - February 12, 2021)
3. In-depth Review Phase (February 13-21, 2021)

Bidding Phase (January 16-22)

Authors who are planning to submit a research artifact are requested to register their artifacts by January 15, 2021 using EasyChair. The submission includes a research abstract with all relevant information and / or links to the repositories containing the information (such as the artifact itself). In exceptional cases described above, the artifact itself may also be submitted as a zip through EasyChair. For more details, please see the submission process described in Section 2.

Immediately after the submission deadline, we will invite you to submit your bids in the EasyChair tool.

The bidding deadline is January 22, 2021.

Please consider your conflicts of interest, your research topics, and your experiences with specific tools and technologies (if applicable) when placing your bids.

Initial Review and Rebuttal Phase (January 23-February 12)

Authors will submit their artifacts by January 22, 2021. We will then assign artifacts to reviewers as soon as possible.

Before the actual in-depth review phase (where no interaction with the authors will take place anymore), reviewers will be asked to check the integrity of the research artifacts and to look for possible setup problems or other smaller technical issues that may prevent the artifact from being properly evaluated (e.g., corrupted or missing files, provided VMs won't start, immediate crashes on the simplest example). During this phase, PC members may contact the authors to request clarifications on the basic installations and start-up procedures or to resolve simple installation problems. Reviewers who wish to communicate with the authors of the artifacts are asked to email the track chairs at artificevaluationicse2021@easychair.org.

In this case we will send the authors and the reviewers a URL to access a chat allowing them to **communicate anonymously** during the rebuttal period. The tool we will use for the communication during the Initial Review and Rebuttal Phase is Etherpad. The orchestration of the communication is done by the PC chairs.

To expedite the review process, we are encouraging the reviewers to try to send all their issues related to installation in *one short message, if possible*. Given the short review time available, the **authors are expected to respond within a 48-hour period**.

Note that we plan to make any communication between a reviewer and the authors visible to other reviewers assigned to the same artifact to mitigate unnecessary overlaps in effort.

The rebuttal period will end on February 12, 2021.

In-depth Review Phase (February 13-21)

After the first quick checks during the initial review and rebuttal phase, possibly leading to the fixing of problems or clarifications during the initial review and rebuttal phase, the actual in-depth review will start. We will use a single-blind review process.

Reviewers review the artifact documentation provided by the authors (e.g. referring to the README file in a repository). Section 2 provides further details about the expected outline of the research artifacts. Except for exceptional cases, the files entailed by the artifact and described in the abstract are already publicly accessible through a repository. In exceptional cases, however, authors might have submitted the files as a package (e.g. zip) through EasyChair: those cases refer primarily to the cases where authors apply for *Reusable* only and where public disclosure of the artifact is not possible, e.g. due to NDAs.

The authors explain in their submission which badges they are aiming for (STATUS file). The reviewers are then asked to review the artifact for the respective criteria (see Section 4) and decide whether the envisioned badge(s) can be awarded, whether an alternative badge should be awarded (provided the submission meets the criteria), or whether no badge can be awarded at all.

Reviewers are expected to assess if and how the things described in the abstract submission are reflected by the actual artifact in the repository. However, we would like to stress the importance to avoid a black and white decision or searching for small issues that prevent issuing a badge. The whole point of this track is to **promote open science in our research community and help authors willing to share their artifacts in doing this correctly (and efficiently)**.

Reviewers are expected to enter the badge decision on EasyChair together with a short review explaining the badge decision. Please note that we do not expect an in-depth review report, but only a short explanation why or why not a certain badge should be awarded. Further note that a paper can receive multiple badges.

The following scores are available in EasyChair to indicate your badge decision:

- NO BADGE: the research artifact does not justify a badge.
- REUSABLE: the research artifact justifies the badge *Reusable*.
- AVAILABLE: the research artifact justifies the badge *Available*.
- REPLICATED: The research artifact is available and justifies the badge *Replicated*.

- **REPRODUCED:** The research artifact is available and justifies the badge *Reproduced*.
- **HIGHER BADGE COMBINATION:** The research artifact deserves a more complex combination of badges for the paper in the form [*Reusable* OR *Available*] AND [*Reproduced* OR *Replicated*]. We do not distinguish further combinations (e.g. artifacts being *Available* are by definition also *Reusable*). Note that for the badges *Replicated* and *Reproduced*, authors will need to offer appropriate documentation that their artifacts have reached that stage.

Note that reviewers are asked to submit their reviews as soon as possible and not to submit all their reviews at once at the end of the review phase. We allow discussions between reviewers to take place at any time during the review phase and all **reviews are made visible to all reviewers of the same artifact as soon as they are submitted** to facilitate effective discussions (and feedback/support by other reviewers) and, again, to mitigate unnecessary overlaps in effort (e.g. to allow reviewers to concentrate on other submissions first).

Finally, it is allowed to involve an external reviewer in cases the reviewer would like to obtain additional feedback or expertise. In that case, it is important to stress the confidentiality of the process to the external reviewer. However, reviewers are expected to also familiarize themselves with the research artifact such that they can assess it fairly. Regardless of the eventual involvement of external reviewers, please note that the PC members assigned to the artifact *are personally responsible for the reviews (with respect to their fairness and accuracy of the decision)*! Furthermore, we expect the PC members to personally participate in the online discussion.

Nominations: If you want to nominate a research artifact (replicated or reproduced) for the best artifact award, please do so by marking it in the review form.

The deadline for submitting reviews is February 21, 2021. *Authors will be notified about the decision on February 24, 2021.*

Summary of Important Dates

The timeline for the artifact evaluation track is as follows:

December 17, 2020: ICSE technical paper notification
 January 15, 2021: Artifact pre-submission registration deadline
 January 22, 2021: AE bidding deadline
 January 22, 2021: Artifact submission deadline
 February 12, 2021: ICSE camera-ready deadline
 February 12, 2021: End of initial review and rebuttal period
 February 21, 2021: AE review submission deadline
 February 24, 2021: Artifact notification

The AE notification is only 12 days after the camera ready deadline for the main research track. It is, thus, essential to stick with this schedule!

4. Evaluation Criteria

The subsequent checklist comprehends a non-exhaustive list of criteria for the evaluation of the artifact submissions for eligibility of the respective badges. We distinguish minimum criteria (which must be met to merit receiving the badge) and optional criteria which we recommend, but do not impose yet as imperative.

Functional and Reusable badge Criteria

For the sake of simplicity, we consider *reusable* as an extension of *functional*. That is, artifacts which qualify for *Reusable*, are per definition *Functional* but not necessarily vice-versa. In any case, as the scope of the AE track is to foster reusability of artifacts (and beyond), we decided to not evaluate and reward *Functional* badges.

Minimum Criteria

- ☑ Artifacts are well documented and offer, at minimum, an inventory of the contents and sufficient description to enable the artifacts to be exercised.
- ☑ Artifacts are relevant to the associated paper and contribute to the generation of its main results.
- ☑ Artifacts are self-contained and exercisable and include scripts and/or software used to generate the results described in the associated paper, i.e. their integrity allows for a successful execution (if applicable, i.e. software-related) and included data can be accessed and appropriately manipulated.
- ☑ Artifacts have a proper licence available for the artifact, explicitly documented in a separate file (e.g. "LICENCE.md"). *
- ☑ Installation Packages have an explicit documentation of the requirements/prerequisites necessary for potential installations or executions of code (e.g. in a file "REQUIREMENTS.md"). Note that this also includes requirements towards operating systems and hardware.
- ☑ Installation Packages have an installation script and step-by-step instructions that allow for the automatic installation of necessary tools and environments. When required environments or operating systems deviate from the norm (which is essentially always the case as there is no real norm), the package must include as well virtual environments (e.g. Docker container image or VirtualBox VM image). The installation must be executable without problems. **

Optional Criteria

- ☑ Artifacts have an indication of the time needed to run them (e.g., 1 hour, 4 hours, 2 days) and how to run a shorter version (e.g., 10 min.) to check that it is functional.

Remarks

* The licence should indicate to the underlying licence model (e.g. creative commons or MIT) and potential restrictions. The licence text should further be self-contained (e.g. by adding the licence text as proposed by, for example, CC BY to the LICENCE.md file). For software, we encourage the use of any open source licence or a Creative Commons licence. For data,

we recommend a Creative Commons licence. In any case, the licence should allow the scientific reuse.

****** Please note that it is the responsibility of submitting authors to provide an installation package that allows to run the artifact in the evaluator's environment. The instructions themselves should be kept to the absolutely required minimum and we recommend relying on virtual environments / automation as much as possible. If the submission includes a simple package with textual files only (e.g. PDFs or spreadsheets), then these documents can be archived in a single package (e.g. zip or tar.gz). The underlying assumption is that if artifacts cannot be installed/exercised without reasonable technical knowledge or without expertise in the research field, then other authors who would make use of that artifact may run into problems as well. In this case, we argue, the badge should not be awarded.

In any case, the identification of potential causes for failed installations or executions is not part of the reviewers' tasks.

Available badge Criteria

The badge for *Available* artifacts extends the *Reusable* badge insofar that the artifact must be made permanently available, i.e. it is publicly available through a preserved, publicly accessible repository with a stable URL and a DOI. In rather rare occasions only, some artifacts may be *Reusable* but still not publicly (permanently) available in that sense (e.g. industry data underlying strict NDAs and, thus, only available upon request to the original authors or artifacts made available through non-persistent repositories). In those cases, the submission of the artifact for review may be done directly through EasyChair.

Minimum Criteria

☒ Previously listed criteria and in addition:

- ☒ Artifact is available for public download from a repository without the need to register.
- ☒ Artifact is available for public download from a persistent repository with a stable URL.***
- ☒ Artifact is associated with a Digital Object Identifier (DOI).

Optional Criteria

- ☒ Artifacts have an explicit documentation of the authors of the artifacts and, ideally, indicators on how to cite them when making use of the artifacts. The authors lists are directly accessible from the main description of the artifact or available through a dedicated file (e.g. "AUTHORS.md").

Remarks

******* We consider temporary drives (e.g. dropbox, google) to be non-persistent, same as individual/institutional websites of the submitting authors, as these are prone to changes. Although not limited to, we strongly recommend relying on services like Zenodo to archiving repositories / repository releases (e.g. from GitHub) as these services are persistent and they also offer the possibility to assign a DOI. In principle, however, publisher repositories (e.g. ACM Digital Library) and open commercial repositories (e.g. figshare) are acceptable as well as long as they offer a declared plan to enable permanent accessibility.

Replicated and Reproduced badge Criteria

The criteria for the replicated and the reproduced badge are primarily assessed based on the submitted research abstracts that outline that (and how) selected artifacts have reached that stage. That is, reviewers are not expected to review the actual reproduction entirely and we expect the abstracts to show that:

- [REPLICATED] the main results of the paper have been obtained in a subsequent study by a person or team other than the original author, using, in part, the artifacts provided by the author.
- [REPRODUCED] the main results of the paper have been independently obtained in a subsequent study by a person or team other than the original authors, *without the use of author-supplied artifacts*.

The main difference between *Replicated* and *Reproduced* lies, therefore, in whether the external replication (partially) needs to rely on artifacts by the authors of the research being replicated or whether the reproduction can be achieved completely independently.

Minimum Criteria

- ☑ The paper reporting on the replication/reproduction has been peer-reviewed.
- ☑ The original paper being reproduced and potentially awarded the badge is publicly available (via a submitted URL directory).
- ☑ Authorships of the reproduced/replicated artifact must not overlap with the reproducing/replicating artifact.
- ☑ The abstract clearly outlines WHAT is being reproduced, WHY it is important, and HOW exactly it has been done. If the replication/reproduction was only partial, then the authors clearly explain what parts could be achieved or which are missing.
- ☑ Submission lays out substantial evidence on replication/reproduction.
- ☑ [For *Reproduced* only] The abstract clearly shows that the main results of the paper have been obtained without author-supplied artifacts.

Optional Criteria

- ☑ Authors pay due respect to the other work related to the reproduction/replication. That is, the abstract is not necessarily critical towards others in the research community.
- ☑ [Mostly only in case the submitting authors are not the ones of the original work being reproduced/replicated but authors nominating original work] Authors provide a critical reflection upon what aspects made it easier/harder to replicate/reproduce and what are the lessons learned from this work that would enable more replication/reproduction in the future for other kinds of tasks or other kinds of research.

Remarks

Please note that to merit the badge *Replicated* or *Reproduced*, it is sufficient if the results are within a margin / tolerance and slightly deviate from those results of the original study as long as the main claims in the original paper are not changed. This is especially true for non-computational studies (e.g., qualitative studies). Also note that it is not the responsibility of the reviewers to completely replicate/reproduce the study by themselves but of the authors to reasonably convey how this has been achieved. The goal of the AE track is to promote work that allows the broader community to use the artifacts, not in-house specialists

only. In case of *Reusable* artifacts emerging from, inter alia, more restrictive industrial research environments, the abstract needs to contain more than unreproducible claims of the artifact being used, i.e. sufficient details on the actual reproduction/replication to convince the well-intended reviewers.

Finally, artifacts may deserve a more complex combination of badges, such as *Reusable* and *Replicated*. For the badges *Replicated* and *Reproduced*, authors will need to offer appropriate documentation that their artifacts have reached that stage.

5. Further supplementary material to this document

While there are various (valuable) contributions related to open science and, thus, related to this guideline, we recommend the following supplementary material. Note that the guideline at hands is intended to be self-contained and the supplementary material is dedicated to the reader interested in the general notion of open science.

A broader introduction into the general notion of open science in Software Engineering, in particular open data and open source which we consider particularly important to the Artifact Evaluation Track, can be found in the (open access) book chapter ‘Open Science in Software Engineering’, available here: <https://doi.org/fjx4>. This chapter contains the ABC of open science and pragmatic, short insights into relevant basics such as proper licensing models.

The recommendations provided in the chapter are also reflected in the ICSE Open Science policy which we recommend to both reviewers and authors alike participating in the artifact evaluation track. Details on the policy can be found here: <https://conf.researchr.org/track/icse-2021/icse-2021-open-science-policies> (and their latest version is available <https://github.com/acmsigsoft/open-science-policies>).

Finally, we recommend the general checklist elaborated by the empirical software engineering research community as the ACM SIGSOFT Empirical Standards for researchers, peer reviewers, editors and publications venues. We refer, in particular, to the supplementary section on open science. The document can be found here: <https://github.com/acmsigsoft/EmpiricalStandards>

Acknowledgements

We would like to thank Paul Grünbacher and Baishakhi Ray (ICSE 2019 Artifact Evaluation Co-chairs) for a previous version of a review process outlined in this guideline. We also want to thank Tim Menzies for his valuable suggestions and collegial advice as well as Michael Dorner, Alessio Ferrari, Davide Fucci, and Daniel Graziotin for their valuable feedback and suggestions on earlier versions of this document.