FAIR4EDTECH: Towards an Artifact Evaluation Track

Jan Bernoth [1], Birte Heinemann [2], Matthias Ehlenz [2], Martin Armbruster [3], and Ulrike Lucke [1]

Abstract: This poster proposes a community-driven Artifact Evaluation Track for the DELFI conference to promote FAIR principles and recognize research artifacts such as software and data. Building on recent initiatives within the AK Open Science of the section Educational Technology of the German Informatics Society, we outline requirements and processes for submitting, evaluating, and awarding badges to artifacts. Emphasis is placed on domain-specific standards in Educational Technology, including suitable metadata ontologies and data exchange formats like xAPI. Review procedures should be time-bound, criteria-based, and ideally include interactive rebuttal phases. The goal is to foster transparency, reproducibility, and reusability in educational technology research by aligning with practices from established computer science venues. The poster invites discussion on how to implement and sustain such a track within DELFI.

Keywords: FAIR Principles, Open Science, Software Evaluation, Metadata Standards, Educational Technology, Artifact Evaluation

1 Motivation

The topic of Open Science, as well as FAIR data and software, has become increasingly important within the DELFI community. Building on a recent position paper, several key challenges were identified: recognizing and crediting software and infrastructure as valuable scientific contributions; improving and diversifying publication practices for software; establishing publication standards that reflect development processes; and fostering reuse and collaboration through open science principles aligned with the FAIR paradigm [KS22]. The position paper led to the formation of a new Working Group (AK Open Science)[4] as part of the section Educational Technologies (FG Bildungstechnologien) within the German Informatics Society (GI). Both workshops [KS23; KS24] of the group were centered on discussions about requirements of publication of research artifacts, conception of peer-review process, and shaping new publication formats. In summary, wishes were that (1) artifact evaluations are divided between research data and research software evaluation, (2) having a more interactive rebuttal phase between reviewers and authors, (3) align reviewing criteria

[1] Universität Potsdam, Institut für Informatik, An der Bahn 2, 14476 Potsdam, Germany,
jan.bernoth@uni-potsdam.de, https://orcid.org/0000-0002-4127-0053;
ulrike.lucke@uni-potsdam.de, https://orcid.org/0000-0003-4049-8088
[2] RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany,
heinemann@cs.rwth-aachen.de, https://orcid.org/0000-0002-7568-0704;
ehlenz@lbz.rwth-aachen.de, https://orcid.org/0000-0001-6189-6056
[3] Karlsruhe Institute of Technology, Am Fasanengarten 5, 76131 Karlsruhe,
martin.armbruster@kit.edu, https://orcid.org/0000-0002-2554-4501
[4] https://fg-bildungstechnologien.gi.de/fachgruppe/arbeitskreise/ak-open-science

doi:10.18420/delfi2025_23

to other software oriented conferences, and (4) present the best artifacts in a conference track.

It is a clear mission that artifact evaluation differs from paper reviewing. Designing research software and evaluating its impact in teaching and learning scenarios is an essential part of educational technology. The research artifacts could differ based on the research question [St24], but mostly the focus is on the educational technology and its impact on the teaching and learning scenario. Due to this, the authors focus in this work on the artifact evaluation track on evaluating software, and expect that most of the measurements could support most of the research data evaluation.

Artifact evaluation tracks are already established in computer science conferences (e. g., the SuperComputing[5] or International Conference on Software Engineering[6]). In these tracks, authors of accepted papers submit the accompanying artifacts for a separate peer review of the artifacts [Ar25]. In addition, authors apply for specific badges, for instance, of the ACM badges [As20]. During the review, reviewers and authors can communicate to clarify upcoming issues or questions. Based on the final reviews, the badges can be awarded for the artifacts.

The goal of this poster is to spark a discussion around community-driven solutions for implementing the FAIR principles within the DELFI community. The focus lies on the idea to establish a dedicated Artifact Evaluation Track as part of the DELFI conference, aiming to make research outputs such as software and data more visible, reusable, and systematically evaluated. Based on the already named preliminary works, the poster presents concrete requirements, formats, and evaluation criteria to enable transparent and fair review processes for research artifacts. Key elements for discussion include the separation of software and data artifacts, the introduction of interactive rebuttal phases, and the alignment with established practices from the software engineering community.

2 Community Solutions for Artifact Creators

Submission to the evaluation track requires the work to be at least FAIR. The recommendations in FAIR4RS [Ch22] are addressed not only to creators of artifacts but also give the research community the task of defining guidelines to achieve FAIR compliance. In the following list, principles from these recommendations, which need to be supported by community solutions, will be cited and possible solutions will be proposed:

1. *F2 – Software us described with rich metadata.* Rich metadata are important for creating findable research artifacts, and using a controlled vocabulary supports this purpose by ensuring consistency and clarity. To provide a broader description of the field to which the artifact belongs, the Computer Science Ontology[7] [Sa18] and

[5] https://sc25.supercomputing.org/program/papers/reproducibility-initiative/
[6] https://conf.researchr.org/track/icse-2025/icse-2025-artifact-evaluation
[7] https://cso.kmi.open.ac.uk/home

the Software Ontology[8] [Ma14] can be used, but these will not deeply describe the purpose. More fitting metadata can be produced by using an ontology that is tied to educational technologies. This compendium about didactic metadata [OR24] provides a good overview of current work on metadata in this field.

2. *I1 – Software reads, writes and exchanges data in a way that meets domain-relevant community standards.* In Education Technology, xAPI is commonly used to log data about learners and teachers in educational systems, although the flexibility of this standard hinders compatibility between systems [Eh20]. Using a centralized registry makes it easier [He22]. For this purpose, we propose to use and, if necessary, to extend the xAPI Definitions Registry[9]

3. *R3 – Software meets domain-relevant community standards.* This principle is also connected with *R1 – Software is described with a plurality of accurate and relevant attributes.* There are three areas that have domain-relevant standards: software or package registries, licenses, and coding practices. For registries, there is a clear recommendation to use language-relevant registries, for example, for Python use PyPI, for JavaScript use npm, or for Unity use Unity's Package Manager. For licenses, we propose to use open licenses, for example, MIT or GNU GPLv3. For coding standards, it is very difficult for such a diverse community to create a standard. A minimal baseline could be created from the guidelines from Lee et al. [Le21], by e. g. using version control, documenting your software and how to set up, and testing mostly at the unit level.

Once these standards are set, the authors can prepare their artifacts for the review phase based on them. For preparing the artifacts themselves, general recommendations like Five Recommendations for FAIR Software [eS] can be used. To put different artifacts into one submission and prepare an overview paper about the artifact submission, the creation workflow of the Research Data Management Container can be used [Be25], a demo can be seen here[10]. The submitters need to clarify for which badge they want to achieve, if the community agrees on ACM Badges [As20] these can be used to choose from.

3 Proposed Process for an Artifact Evaluation Track

For reviewing an artifact, there should be an explicitly mentioned expected time frame, for example, two hours, so that reviewers know the required time investment ahead and authors can prepare their artifact accordingly. As a consequence, reviewers need to efficiently and effectively evaluate and check the quality criteria corresponding to the badges for which the authors apply. They can encompass, for instance, the metadata, data formats for data exchange, or, inspired by the ACM badges [As20], if the artifacts are documented, consistent,

[8] https://github.com/allysonlister/swo
[9] https://xapi.elearn.rwth-aachen.de/
[10] https://rdmc.create.nfdixcs.org/

complete, and exercisable. Regarding “exercisable”, reviewers should execute the artifact to check the results with the associated paper or claims. To simplify reviewing, all checks should be automated as much as possible.

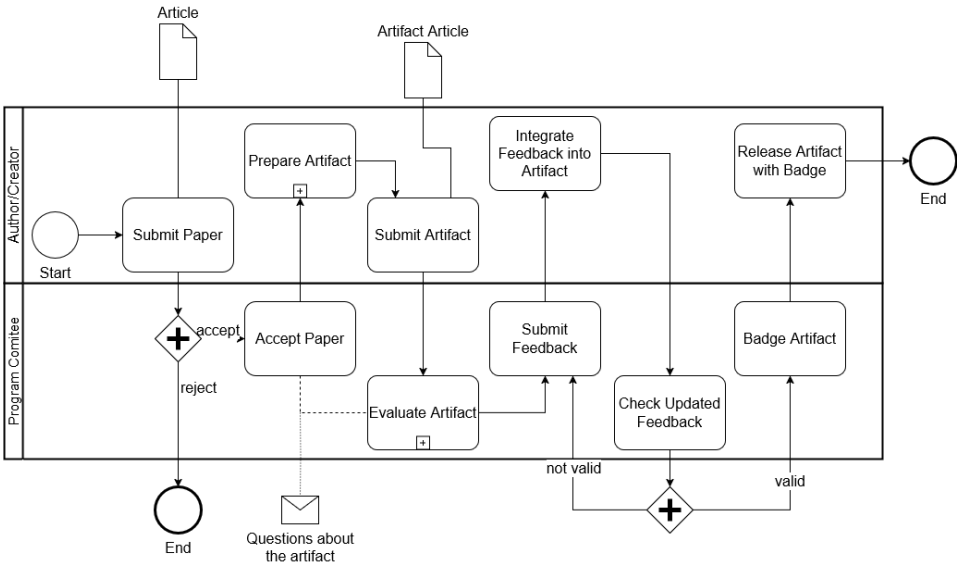

Fig. 1: Proposed Workflow for the Artifact Evaluation in BPMN 2.0

In Figure 1, a possible process of the Artifact Evaluation Track is proposed. The authors should indicate their intent to submit an artifact for evaluation, to be provided upon paper acceptance. After paper acceptance, the authors should get time to prepare their artifact. Some conferences also have a communication phase in the evaluation period in which reviewers can communicate with the authors to quickly resolve issues and have a running system in their environment. After the artifacts are evaluated, the committee of reviewers may award the badge. The authors need to create a new release of their artifacts with eventual updates from the review phase and the awarded badge.

In summary, this poster proposes a community-driven Artifact Evaluation Track for the DELFI conference to promote FAIR principles and recognize research artifacts like software and data. Building on prior Open Science initiatives and related efforts, it outlines requirements, like metadata standards and review processes, including interactive rebuttals and badge systems to foster transparency, reproducibility, and reusability.

Acknowledgments

This work was mainly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 52/1 – 501930651.

References

[Ar25] Armbruster, M.: Artifact Evaluation in the Context of NFDIxCS, 2025, https://nfdixcs.org/meldung/artifact-evaluation-in-the-context-of-nfdixcs.

[As20] Association for Computing Machinery: Artifact Review and Badging Version 1.1, 2020, https://www.acm.org/publications/policies/artifact-review-and-badging-current, accessed: 03/30/2023.

[Be25] Bernoth, J. et al.: Workflow for Creating and Sealing a Research Data Management Container (RDMC). In (Feichtinger, K.; Sonnleithner. Lisa; Hajiabadi, H., eds.): Software Engineering 2025 – Companion Proceedings. Gesellschaft für Informatik e.V, Bonn, 2025.

[Ch22] Chue Hong, N. P. et al.: FAIR Principles for Research Software (FAIR4RS Principles), 2022.

[Eh20] Ehlenz, M. et al.: Eine forschungspraktische Perspektive auf xAPI-Registries. In: DELFI 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V. Gesellschaft für Informatik e.V., Bonn, 2020, http://dl.gi.de/handle/20.500.12116/34179.

[eS] eScience Center Netherlands: Five Recommendations for FAIR Software, https://www.fair-software.eu/, accessed: 06/06/2025.

[He22] Heinemann, B. et al.: xAPI Made Easy: A Learning Analytics Infrastructure for Interdisciplinary Projects. International Journal of Online and Biomedical Engineering (iJOE) 18 (14), pp. 99–113, 2022.

[KS22] Kiesler, N.; Schiffner, D.: On the Lack of Recognition of Software Artifacts and IT Infrastructure in Educational Technology Research. In: 20. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V, Bonn, pp. 201–206, 2022.

[KS23] Kiesler, N.; Schiffner, D.: Open Science in den Bildungstechnologien: Zur Publikation und Begutachtung von Forschungsdaten inklusive Software im Rahmen der DELFI. In: Workshops der 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V, Bonn, pp. 159–168, 2023.

[KS24] Kiesler, N.; Schiffner, D.: Open Science in den Bildungstechnologien: Zur Unterstützung bei der Veröffentlichung von Forschungsdaten. In (Kiesler, N.; Schulz, S., eds.): Proceedings of DELFI Workshops 2024. Gesellschaft für Informatik e. V. (GI), Bonn, pp. 77–86, 2024.

[Le21] Lee, G. et al.: Barely sufficient practices in scientific computing. Patterns (New York, N.Y.) 2 (2), p. 100206, 2021.

[Ma14] Malone, J. et al.: The Software Ontology (SWO): a resource for reproducibility in biomedical data analysis, curation and digital preservation. Journal of biomedical semantics 5, p. 25, 2014.

[OR24] Oellers, M.; Rörtgen, S.: Kompendium: Didaktische Metadaten, 2024.

[Sa18] Salatino, A. A. et al.: The Computer Science Ontology: A Large-Scale Taxonomy of Research Areas. In (Vrandečić, D. et al., eds.): The Semantic Web – ISWC 2018. Vol. 11137, Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 187–205, 2018.

[St24] Striewe, M.: Forschungsdaten in der Bildungstechnologie: Worüber sprechen wir und was ist wann relevant?, 2024.