An Experience Report on Artifact Evaluation in Brazilian Conferences

Tiago Heinrich
theinric@mpi-inf.mpg.de
Max Planck Institute for Informatics (MPI)
Saarbrücken, Saarland
Germany

Diego Kreutz
diegokreutz@-unipampa.edu.br
Federal University of Pampa (UNIPAMPA)
Alegrete, Rio Grande do Sul
Brazil

Roben Lunardi
roben.lunardi@-restinga.ifrs.edu.br
Federal Institute of Rio Grande do Sul (IFRS)
Restinga, Rio Grande do Sul
Brazil

Rodrigo Mansilha
mansilha@unipampa.edu.br
Federal University of Pampa (UNIPAMPA)
Alegrete, Rio Grande do Sul
Brazil

Vinicius Fulber-Garcia
vinicius@inf.ufpr.br
Federal University of Paraná (UFPR)
Curitiba, Paraná, Brazil

Lourenco Alves Pereira Jr
lourenco.junior@gp.ita.br
Aeronautics Institute of Technology (ITA)
São José dos Campos, São Paulo, Brazil

Rafael R. Obelheiro
rafael.obelheiro@udesc.br
State University of Santa Catarina (UDESC)
Joinville, Santa Catarina
Brazil

This article is an editorial note submitted to CCR. It has NOT been peer reviewed.
The authors take full responsibility for this article's technical content. Comments can be posted through CCR Online.

ABSTRACT

An artifact is a resource essential to the complete understanding of a scientific article, extending beyond the manuscript itself. As more conferences allow authors to submit artifacts alongside their articles, different review processes, badge types are emerging, and badge-acquisition requirements. In this paper, we describe the process for conducting artifact evaluation that has been implemented in the Brazilian community. Our discussion focuses on the two largest Brazilian networking and security venues, where we discuss the challenges we encountered during our interactions and the solutions we ultimately adopted to improve the review process.

CCS CONCEPTS

• **General and reference** → **Evaluation**; • **Software and its engineering** → **Reusability**; *Software configuration management and version control systems*; **Software post-development issues**; *Extra-functional properties*; • **Social and professional topics** → *Software management*;

KEYWORDS

Research Artifacts, Artifact Evaluations

1 INTRODUCTION

The most popular way of disseminating scientific results today is through publishing articles in journals, conference proceedings, and open databases. However, the complexity of proposals and experiments may not be fully captured in a manuscript alone, requiring additional materials to meet the scientific method requirements; for example, to ensure reproducibility and falsifiability. A way to achieve that is to provide artifacts alongside the articles. Artifacts, in turn, can be defined as digital objects created by the authors of an article or generated from the experiments they conduct [4].

In this way, it is reasonable to say that published articles go beyond the manuscript itself, with an artifact being not merely an auxiliary resource but an essential element for fully understanding the presented proposal [5, 8, 9]. These artifacts ultimately enable other researchers to reproduce and verify results, build upon previous findings, and extend the frontiers of knowledge by straightforwardly using publicly available resources [1, 3].

To evaluate and recognize artifacts, the Artifact Evaluation Committee (AEC) was introduced over a decade ago, first presented in the software engineering community at ESEC/FSE 2011 [2, 6]. The process of evaluating an artifact mainly consists of the validation of the artifact in three aspects[1], these being: (i) **availability**, the content is made available for retrieval, permanently and publicly; (ii) **functionality**, compliance with the expectations set by the article in terms of functionality, usability, and relevance; and (iii) **reproducibility**, can be used to reproduce the main results presented in the article.

Over the following years, the importance of the artifacts and AEC initiatives spread to multiple communities, including ACM SIGCOMM, NSDI, EuroSys, and USENIX, which addressed them at their academic conferences. Inspired by these initiatives, in 2022, members of the Brazilian networking and cybersecurity research communities began a broad discussion on how local events could promote artifacts and incorporate their evaluation into the traditional peer-review process for scientific articles. Before, conferences in Brazil did not offer formal mechanisms for submitting and evaluating anything beyond the articles themselves. Breaking with this habit, over the course of 3 years, we have been implementing and

[1] It may vary according to the evaluation policy in place.

iteratively refining a process for popularizing and evaluating artifacts within the Brazilian Computing Society, primarily in the networking and cybersecurity communities.

The described process was initially executed in 2023 in the Brazilian Computer Networking and Distributed Systems Symposium (SBRC)[2] and the Brazilian Symposium on Cybersecurity (SBSeg)[3]. SBRC is the leading Brazilian conference on networking and distributed systems, with 43 editions since 1980. In 2025, the conference accepted 74 full papers (out of 203 submissions) with 16 artifacts accepted, and had about 600 attendees. SBSeg is the top Brazilian conference on cybersecurity: it has been held annually since 2001, and the 2025 edition featured 60 full papers (out of 163 submissions) and 24 short papers (out of 66 submissions), with 38 artifacts accepted and an attendance of 370 people. Both conferences are annually organized by the Brazilian Computer Society (SBC), the largest computer society in Latin America.

Given the scenario described, this paper presents the measures we took, the challenges encountered, and the lessons learned, leading to the current state of the artifact evaluation process in Brazil. It contributes by (i) documenting the design and evolution of an artifact evaluation process tailored to a national research community, (ii) proposing a badge model that extends existing ACM-based policies with sustainability-oriented criteria, and (iii) reporting empirical evidence of process maturation and artifact quality improvement over multiple conference editions. The remainder of the article is organized as follows: Section 2 presents the review process before we establish an AEC; Section 3 presents our policies and badges; Section 4 presents our guidelines; Section 5 explains our review process; Section 6 discuss the lessons learned; and Section 7 concludes the work.

2 THE PROCESS BEFORE THE AEC

Historically, Brazilian conferences have not maintained a formal process for artifact review. Consequently, when authors wished to submit an artifact along with their paper, it was merely referenced within the manuscript, without a defined format, citation requirements, or minimum criteria to ensure the artifact's usability, and hosted on any chosen platform, with no established review procedure. Neither workshop nor main track submissions required artifact submission, nor did they include a review process for such materials.

A key innovation came from the Tool Showcase track, which mandated the submission of artifacts alongside papers, a requirement central to it is focus on tool development and implementation. This track was first introduced in 2001 at SBRC and in 2017 at SBSeg, and it has been included ever since. Once a paper is accepted to the Tool Showcase track, authors must include a link to the corresponding artifacts within the paper. Notably, the only criterion evaluated by the Technical Program Committee for this aspect is the presence or absence of this link, alongside all the regular scientific criteria (e.g., novelty, impact).

Therefore, the previous artifact submission and review process did not consider criteria such as availability, documentation, functionality, and reproducibility. Due to the limitations encountered in this previous process, the quality of the artifacts presented some recurring problems, such as (i) missing or a lack of proper documentation to understand and execute the artifact; (ii) no real standardization in the artifact link presented in the article, difficulties in locating the resource; (iii) non functional artifacts since an installation from scratch has never been carried out considering different environments; and (iv) reproduction of values found in the article ends up not being taken into consideration in the shared material, with missing resources and omitted information.

We recognize that the quality of artifacts is as essential as that of the article itself. Artifacts play a fundamental role in the complete understanding of a scientific article, adding a practical dimension that enriches the academic content. Having an established artifact evaluation process plays a crucial role in enhancing the quality of the artifacts themselves, for example, by clearly defining the expected requirements for an artifact.

The new process for reviewing and submitting artifacts that we began to design drew inspiration from the processes being carried out by SIGCOMM, USENIX, NSDI, and EuroSys. Also, it took into consideration the problems we observed in the existing processes adopted in SBRC and SBSeg. The new process should raise awareness among authors and reviewers about the available badges and their expected requirements. The goal was not only to increase the quality of the submitted papers and artifacts, but also to contribute to the continuous improvement of the scientific community as a whole. In the initial stages, we implemented the artifact review process only for the Tool Showcase track; later, the process was adopted in the SBSeg main track. Currently, track and workshop chairs can opt for incorporating artifact evaluation.

Similar challenges were reported in early artifact initiatives in international venues prior to the formalization of AECs, particularly regarding missing documentation, non-functional code, and lack of reproducibility guaranties. The Brazilian context mirrored these issues, reinforcing the need for a structured and enforceable evaluation process.

3 ARTIFACT BADGES AND POLICY

Conferences that have adopted an artifact evaluation process usually follow a variation from the ACM artifact policy first presented in 2017 and revised in 2020 (ACM Badging V1.1) [1, 7]. This policy specifies three badges that must be judged during a review process: artifacts available, artifacts evaluated (also known as functional or reusable), and results validated (also known as reproduced or replicated). This policy does not define review procedures or framework, providing a broad definition that allows the proposed badges to have a general comparability across communities. ACM establishes the following definition for each badge:

Artifacts Available: the artifact must be publicly accessible in an archival repository. A DOI or link to this repository, along with a unique identifier for the object, is provided;

Artifacts Functional: the artifact must be documented, consistent, complete, exercisable, and include appropriate evidence of verification and validation; and

Results Reproduced: the artifact must enable a person or team other than the author to achieve the same results as the ones presented in the article.

[2] https://sol.sbc.org.br/index.php/sbrc/issue/archive
[3] https://sol.sbc.org.br/index.php/sbseg/issue/archive

Some conferences adopted these definitions with minor variations. NSDI'26, USENIX'25 and EuroSys'25 used three badges: artifacts available, artifacts functional, and results reproduced. SIGCOMM'25, in turn, employed only two badges: artifacts available and artifacts evaluated/functional. Even though each conference slightly adapted the requirements for awarding badges, they still follow the criteria stipulated by ACM Badging V1.1. Therefore, we will not delve into the differences, and we will only discuss the criteria we established for an artifact to receive a badge. We considered using the same set of badges proposed by SIGCOMM; however, we chose to use four badges, splitting criteria that would be evaluated together in a single badge, and we also opted to evaluate the quality of the software in relation to its sustainability. Therefore, the badge definitions we have used over the last four years are:

Artifact Available: A badge granted to articles that have their associated artifact (code and/or data) available in a stable repository such as GitHub, GitLab, or an institutional platform. In the repository, we expect to find a `README.md` file that approaches at least nine topics: (i) the project title, with a summary describing the artifacts' objectives; (ii) a description of the repository's organizational structure; (iii) the labels considered in the artifacts review process; (iv) a description of the required environment (hardware and software) for running the artifacts (and reproducing the results from paper, if that is the case); (v) the benchmarks used and the dependencies for execution (dependency versions and processes for accessing third-party resources, if necessary); (vi) if running an artifact poses any risk to the reviewer, this risk should be described, and the appropriate process for mitigating the risk should be presented; (vii) the download and installation instructions for the artifacts; (viii) a step-by-step guide for executing a minimal test. A minimal test execution allows reviewers to observe some of the artifacts' functionalities. This test helps identify problems during the installation process. (ix) a step-by-step guide for executing and obtaining the results in the paper. It allows reviewers to achieve the stated results. Each claim should be presented in a subsection, detailing the configuration files to be changed, the commands to be executed, the flags to be used, the expected execution time, the expected resources to be utilized (such as 1GB RAM/disk), and the expected results. If reproducing all experiments is not possible within a reasonable time frame, authors should identify the central claims in the article and give instructions for reproducing them; and (x) the license to which the artifact is subject.

Artifact Functional: The code and/or artifact is expected to be executable, and the reviewer will be able to observe some of its functionality. To acquire this badge, additional information must be included in the repository's `README.md`, such as a list of dependencies; a list of dependencies, languages, and environment versions; a description of the execution environment; installation and execution instructions; a minimal execution example.

Artifact Sustainable: The artifacts, mainly code-based ones, are expected to be modularized, organized, and easy to understand. To obtain this badge, there must be minimal code documentation (describing files, functions, etc.), a minimum level of code readability, and reviewers must be able to identify the article's central claims within the artifact easily.

Artifact Reproducible: The reviewer is expected to be able to reproduce the central claims presented in the article through the artifact. To obtain this badge, the following are expected: instructions for implementing the major claims (*e.g.*, results from the main graphs/tables); and a description of the process by which the experiments were performed to achieve the article's results.

The decision to introduce a separate Artifact Sustainable badge emerged from repeated observations that artifacts could be functional yet difficult to understand, maintain, or extend. By isolating sustainability concerns, the evaluation process encourages authors to invest in code organization and documentation beyond mere executability, addressing long-term reuse and community impact.

The four badges presented, along with their respective requirements, are all considered during the review process. To optimize the evaluation and reading process, we defined dependencies among them: an artifact will only be evaluated on functionality if it has already received the availability badge, and will only be evaluated in terms of sustainability and reproducibility if it has previously received the functionality badge. SIGCOMM asks reviewers to answer questions that are somewhat subjective (e.g., are the artifacts sufficiently documented? do they include all of the key components described in the paper?). We define objective questions (e.g., are the instructions provided by the authors sufficient for running the artifact? can you reproduce the results in the paper?) that establish minimum criteria for awarding a badge; therefore, reviewers should check that artifacts meet these criteria. Along with the definition of the badge, we also present examples of the criteria that reviewers should consider. From the authors' perspective, in addition to receiving the definitions of each badge and instructions on how reviewers will carry out the review process, templates of the minimum criteria (e.g., sections of the `README.md` file) are also published.

4 AUTHOR AND REVIEWER GUIDELINES

The AEC work begins after the Program Committee has decided on the papers submitted for the conference. Authors with accepted papers are encouraged to submit their artifacts. In SBRC and SBSeg, artifact submission is mandatory only for the Tools Showcase track, and optional for other tracks. Authors are given between one and two weeks to prepare the artifacts and submit them for review. To foster transparency about the review process, artifact requirements, and badge requirements, the instructions for authors and reviewers are similar. Artifact review is single-blind.

We highlight to authors that the AEC review process is independent of the paper review, and that the review committee expects that: (i) artifacts comply with the minimal requirements for the requested badges; (ii) any additional information needed is provided in an optional appendix that shall be later published in the repository; (iii) authors respond to the questions posted by reviewers on the platform and submit the rebuttal letter by the deadline; (iv) authors give access to additional resources required to test the artifacts (*e.g.*, specific hardware, cloud, SSH); (v) the authors have already tested the artifacts in a clean environment (*e.g.*, a virtual machine) to avoid installation problems. Unlike other artifact review processes, we make these expectations clear to the authors to streamline the review process.

In cases where specific resources or restrictions apply to an artifact, authors have the option of submitting an appendix with

additional information, such as instructions for accessing cloud infrastructure or SSH keys. Any additional resources are expected to be made available by the authors to the reviewers, and this information should be included in the appendix.

From the reviewers' perspective, the instructions are similar to those provided to the authors, with additional information about the steps and duties of the AEC. Every year, we invite new reviewers with prior experience in artifact evaluation. All reviewers can indicate which artifacts they would like to review through a bidding process (similar to the traditional paper bidding used in TPCs). However, before bidding starts, we hold a kick-off meeting to present the steps of reviewing an artifact, outline how interaction with authors should be carried out, what is expected from reviewers, and which elements in artifacts should be considered for badges. We also discuss some reviews from previous editions and answer questions from reviewers. The kick-off meeting is recorded and shared with all the AEC members. This process has been quite effective and we recommend a similar process for SIGCOMM-sponsored conferences, as we noticed that some reviewers had incorrect expectations (*e.g.*, they expected the AEC process to be similar to the PC). The discussions helped align expectations and provide guidance.

We have used HotCRP[4] for managing the artifact evaluation process. In the first two evaluations, we utilized additional communication channels beyond email and HotCRP (*e.g.*, WhatsApp and Telegram groups) to address reviewers' questions and facilitate discussion. Afterwards, we decided to limit discussions to HotCRP so that reviews are concentrated in a single communication channel.

While the detailed guidelines increase transparency and review quality, they also introduce an initial learning curve for authors unfamiliar with artifact evaluation practices. Our experience shows that this cost is front-loaded and decreases substantially in subsequent submissions, particularly when authors reuse templates and previously validated structures.

5 REVIEW PROCESS

After the bidding period, reviewers are assigned to a set of artifacts, aiming to maximize their demands and affinities (as defined in the bid) with the assigned reviews. Then, the review process begins. Over the years, we tested different review loads as our submissions experienced regular growth. We found that a load of more than two artifacts per reviewer is not ideal, as reviewers are volunteers with limited time to perform reviews. Thus, we strive to minimize any burden during the process, which is why we have established well-defined requirements for each badge and for authors submitting artifacts.

The review process takes, on average, one month, and authors and reviewers can interact freely to improve the artifact so that all badges can be attributed. However, we noticed that a few reviewers initiated the review process within the first two weeks; this ultimately overloaded authors, who had less time to make changes and wait for reviewers' feedback. For this reason, we adapted the review process over the years to improve this interaction.

The first adaptation was to ask reviewers to verify that an artifact is available, all necessary resources are located in the repository, and hardware requirements are reported. This verification is done by the chairs, who message reviewers to confirm that everything is in order for the review to begin.

The second adaptation was made in the review stage itself. As in SIGCOMM, our first three experiences had a single round, allowing authors and reviewers to interact over three weeks. A clear problem we observed on the part of the authors was delaying until the last minute to solve problems raised by reviewers, which left too little time to review the fixed versions. Therefore, we split the AEC review process into more rounds, and the current version consists of:

Review Round 1: Reviewers review artifacts considering the evaluation criteria. During this process, users can post messages on the platform. These can be discussions between review committee members, such as questions for the authors, or questions regarding issues found in the artifact.
At the end of the round, reviewers must submit a report to be presented to the authors. This report should highlight the steps taken to evaluate each badge, problems that were observed and were still unsolved after the discussions, and the results achieved (problems in the implementation process should be clearly explained in the review).
At the end of this process, no decision is presented to the authors; instead, the entire discussion takes place over two weeks, aiming to raise the main problems and allowing the authors to address them through discussions with the reviewers.

Rebuttal Phase: After Round 1, authors have one week to address the problems raised by the reviewers and submit a rebuttal letter explaining the changes made to the artifact. At this stage, authors, based on the review conducted in the first stage, must clarify any doubts, resolve any issues, report any mistakes, and/or explain any aspects that the reviewers overlooked.

Review Round 2: The reviewer considers the points raised in the first stage and the rebuttal phase to verify if the artifact meets the minimal criteria for each badge and makes a decision on which badges should or should not be awarded.

These changes have fostered a higher level of interaction aimed at improving and resolving issues found in artifacts. There is also a reduction in the workload for reviewers, who now seek to inform authors of issues and, through two rounds, ensure that action is taken within a reasonable timeframe before a decision is made. As the process changed, the quality of the artifacts has improved. Table 1 highlights this evolution: while in the first experience only 56.25% of the submitted artifacts were deemed available and a mere 13.04% were assessed as reproducible, in the latest (SBSeg'25) all 38 artifacts received the availability badge and 57.89% received the reproducibility badge,

Compared to the single-round artifact evaluation models adopted in earlier editions of SIGCOMM and similar venues, the multi-round process adopted here reduced last-minute fixes, improved reviewer confidence, and led to more consistent badge decisions, at the cost of slightly increased coordination effort by the chairs.

We experimented with different review forms over the years to gather information that makes the decision from reviewers clear for

[4]https://github.com/kohler/hotcrp

authors and chairs. The main elements of our review form consist of:

- For each badge, we request a description of the process with detailed information about the requirements that are covered/not covered by the artifact.
- Additionally, for functionality and reproducibility badges, another field is present that consists of proof of execution. In this field, we request that reviewers publish the execution logs (*e.g.*, success or error) and reproducibility outcomes (*e.g.*, observed results). This execution field has become essential for detecting problems in the artifact, as it demonstrates to the authors the problems found, allows comparison between reviewers' results, and ultimately serves as the basis for the points raised by the reviewers. We do not accept reviews without a valid proof of execution, as we require reviewers to present detailed information about the execution (this way, misunderstandings are easily identified), the results obtained (which should be comparable between reviewers), and any functional errors that were found.
- To reward the best artifacts, we also added a field where reviewers can select how they think the artifact is ranked. Because reviewers have limited visibility, as they typically review only a small portion of the overall number of artifacts, we provide guidance on what to expect from an artifact in the Top 5%, 25%, 50%, and below. Therefore, reviewers are expected to award higher scores (5% and 25%) to artifacts with at least three badges awarded and that stand out in terms of quality compared to the others. Furthermore, artifacts that have received no more than two awards are expected to receive a score no higher than the minimum.

After the reviews are submitted from the AEC, the results are published, and the authors are notified. Thus, we openly invite members of the AEC and external candidates (typically at least two external members) to participate in the selection of awards. The *selection committee* usually consists of five members (one of the AE chairs, two AEC members, and two external members), who are tasked with defining the rules for selecting the best artifacts and reviewers. They receive the rules stipulated in the previous year and use these rules as a basis for making changes as necessary. In 2025, the rules defined to award papers were: (i) at least three badges awarded; (ii) submission of the rebuttal letter, with required clarifications and necessary modifications to improve the artifact; (iii) at least two reviewers (out of three) ranked the artifact in the Top 5%; and (iv) authors respond to most of the reviewers' questions, aiming to improve the quality of the artifact with feedback incorporated into the artifact (when applicable).

To recognize the best reviewers, the selection committee considered the following rules: (i) submitted proof of execution of the artifacts (description of the reproduction or execution process); (ii) checked that the minimum requirements for each evaluated badge were met; (iii) completed the Artifact Functional and Artifact Reproducible badge evaluation for at least two artifacts; (iv) provided suggestions to the authors for at least two of the evaluated artifacts; (v) engaged in discussions among the reviewers (or authors) of at least one artifact (when applicable).

6 LESSONS LEARNED

Our experience leading AECs over the last four years has shown the importance of this initiative. Artifacts are more than just implementation efforts that demonstrate a research proposal's feasibility; a good artifact is a well-documented, reusable tool that plays a crucial role in making the research reproducible and enabling it to serve as a basis for subsequent development. On the other hand, artifact evaluation is a laborious, time-intensive task; it often takes longer to review an artifact than a paper. Therefore, our main goals as AEC chairs have been to (i) recognize artifact quality and incentivize its improvement and (ii) streamline the process to ensure desirable outcomes while minimizing the burden on reviewers.

We observed a lack of understanding of the process for authors submitting artifacts. Due to unfamiliarity with other artifact submission processes, many authors did not strictly adhere to the established criteria for artifact submission, resulting in the rejection of the artifact or, depending on the track to which the article was ultimately submitted, a rescinded acceptance of the article.

Authors often left responding to comments or clarifying questions until the final days of the review process. Consequently, these reviews, conducted in a limited time window, often presented problems due to the delayed process, since reviewers cannot perform miracles and complete the entire review in one night. The irony is that these were the authors who ended up being most critical of why a badge had not been awarded. Consequently, our review model evolved to minimize these issues, where we define strict time windows during the review process.

With the popularization of the artifact submission process, we have also observed an increase in the number of artifacts submitted each year. For instance, SBSeg went from 29 artifacts (out of 70 accepted papers) in 2024 to 38 artifacts (out of 84 papers) in 2025, a 31.0% increase. Therefore, we have noticed the difficulty in finding qualified reviewers willing to participate in the review process. We have added a larger number of reviewers from the industry and early-stage graduate students in recent years to meet this demand. The training provided during the kick-off meeting proved essential to maintaining the quality of the reviews. The review form and the minimum criteria for a submission are essential for maintaining good quality in the reviews. The "proof of execution" field enables reviewers, especially inexperienced ones, to accurately assess functionality and reproducibility, while also facilitating easy comparison between reviews. Consequently, presenting proof of the artifact's execution allows authors to understand the limitations encountered by reviewers.

Chairs should constantly monitor interactions between authors and reviewers to avoid misunderstandings and streamline the process. Due to the number of artifacts in a review cycle, it is ideal to interact with reviewers who have not yet begun their reviews or who have encountered obstacles during the reproduction of the artifact. These interactions should occur as early as possible in each review cycle. If issues are observed (e.g., a reviewer not being active), assigning a new reviewer as soon as possible is essential to avoid delays in the process. We generally maintain a pool of reviewers who have participated in multiple review cycles with a reduced review load to assist in these situations.

Key lessons learned include: (i) artifact review often requires more effort than paper review; (ii) clear, objective badge criteria significantly reduce reviewer subjectivity; (iii) mandatory proof of execution is essential for assessing reproducibility; and (iv) early and monitored interaction between authors and reviewers is critical to process efficiency.

6.1 Participant Feedback

After each iteration of the process, a feedback form is sent to all participants to improve the review process. Considering the interactions carried out in 2025, we highlight some key points that will be taken into account for the subsequent experiences. Only 6.25% of reviewers thought the review form could be improved in some way. 68.7% of reviewers needed more than 4 hours to review an artifact. 75.0% of reviewers indicate that a review period of two to three weeks is sufficient.

The submission instructions are clear, detailing the criteria that will be used to evaluate each badge. However, it would be helpful to present specific scenarios where badges cannot be awarded or where more information is required from the authors.

For review, authors may submit a virtualized environment where the artifact is already pre-configured. Feedback from the reviewers indicated a preference for submissions that presented a Docker image, due to its ease of reproduction and ease of understanding. We ultimately recommended that authors use such an environment whenever possible.

In the reproduction process of the work, the lack of information and scripts used to generate the tables and figures found in the paper is a common occurrence. This resource is not essential for the artifact, but it brings more convenience to the process of reproducing and validating the observed results.

In addition to hardware details, we also began requesting the average processing time relative to the available hardware. Some submissions, despite being well-documented, may be hampered by reviewers' hardware limitations due to their long processing times. This information helps identify issues early in the review process rather than at the end.

We also asked about the suitability of the HotCRP platform for the artifact review process. In general, reviewers felt there was a lack of a simple way to attach logs and images to the reviewers' files, which would be made available to authors. This information proved crucial in enhancing the quality of the artifacts and is currently shared via an external link, as the form does not permit the addition of such resources.

6.2 Limitations

Ethical considerations are constantly evolving, becoming essential for conducting research. Our submission process still has limitations when it comes to assessing ethical issues that may arise due to the publication of the artifact. Our community has limited access to Institutional Review Boards (IRBs), which ultimately places a greater burden on the reviewers, who, depending on their experience and research areas, may not be qualified to evaluate ethical research criteria.

We rely on the PC review process to determine whether any risks exist in publishing the paper and artifact, with AEC evaluating only the ethical implications of the artifact itself. The AEC evaluates whether sensitive or private data to be published may present a security risk.

We also consider royalties for the publication of an artifact. Although rare, we have had occasions where artifacts came from the industry, and part of the code and data were proprietary. With the requirements for availability, it is impossible to accept such an artifact, which, depending on the track, may result in the article acceptance being rescinded. In this situation, we work with the authors to publish the artifact in a way that minimizes the need for proprietary content.

With the rapid adoption of generative artificial intelligence, we were concerned about the possibility of reviews being generated by such tools. However, due to the steps of the review process and the criteria for review acceptance, we were able to minimize such issues. Nevertheless, this comes with an additional burden for chairs who must periodically interact with reviewers at a detailed level for each artifact.

7 CONCLUSION

We have presented our recent experience with the artifact review process in the Brazilian scenario, particularly within the networking and cybersecurity communities. Through the lessons learned and adaptations made in relation to international references, we defined and presented an artifact review model that is adopted by two of the largest conferences in the computer science community in Brazil, SBRC and SBSeg.

This process has fostered improvements to artifact submission in terms of both quantity and quality. Looking towards the future, we feel that the culture of artifact submission still needs to be consolidated: authors should be encouraged to submit artifacts and have them evaluated, but this should not be mandatory. We noticed that there is a higher overhead in carrying out the artifact review process, due to the need for interaction between authors and reviewers to solve problems, clarify doubts about the documentation, and improve the quality of the artifact, which requires more time for interactions from everyone involved. We can only minimize this overhead by introducing good practices and raising awareness within the community that is making the artifacts available.

We do not believe that requiring authors to submit artifacts is a good practice for events. An artifact needs to have a level of maturity for the community to use it, with documentation at a level of quality that allows its use. Furthermore, when making artifacts available, authors should be conscious of the need to provide all the content/material necessary for the complete reproduction of the artifact. We need to be aware that an artifact is not a stand-alone resource; a lack of maturity in an artifact can reflect problems that are visible in the submitted paper.

As the culture of artifact submission becomes more ingrained, we may devise mechanisms for incentivizing artifacts. Today, the evaluation of artifacts is dissociated from the evaluation of the corresponding papers: an artifact badge is a bonus that does not influence the acceptance or rejection of a paper. We could, for instance, use badges to enhance paper scores before ranking the papers. Issues that require further reflection include:

(1) Which badges would warrant increasing a paper's score, and by how much (e.g., if scores are in a 0–10 scale, earning the Artifact Reproducible badge could add 0.5 to the score); and
(2) We must take care to avoid penalizing papers for which artifacts are not applicable (e.g., theoretical papers).

Over time, one particular concern we have identified is the long-term maintainability of artifacts. To address this issue, our first step was introducing the Artifact Sustainable badge, bringing it to the forefront. In all, having clear badge requirements with detailed documentation (including examples) and constant interaction between reviewers and authors to has helped evolve the artifact review process.

The growing importance and popularity of artifacts will undoubtedly pose new challenges in the coming years, prompting the research community to continually assess and revise its methods to address them. Artifacts facilitate the thorough application of scientific methodology, leading to increasingly effective academic and technical contributions. By presenting our process and the lessons learned during its evolution, we hope to encourage other research communities to embrace artifact evaluation and help them to institutionalize it in a gradual and sustainable manner.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to the Brazilian networking and cybersecurity research communities for embracing artifact evaluation. We thank the SBRC and SBSeg organizers for providing us space, UFPR and C3SL for providing computing resources for running our HotCRP instance and testing artifacts, the authors for submitting their artifacts and providing valuable feedback, and the dedicated work of our reviewers and committee members who helped us improve the review process. Rodrigo Mansilha acknowledges the support of FAPERGS (grant no. 25/2551-0002572-9) and FAPESP (grants no. 2020/05183-0 and 2023/00816-2). Rafael R. Obelheiro acknowledges the financial support from UDESC and FAPESC.

A AWARDED BADGES

Table 1 presents an overview of the badges awarded over the years. Percentages are relative to the number of artifacts submitted in each conference.

	artifacts	available	functional	reproducible	sustainable
SBRC'23	23	56.52%	21.74%	13.04%	8.70%
SBRC'24	10	70.00%	20.00%	10.00%	10.00%
SBRC'25	16	100.00%	87.50%	56.25%	43.75%
SBSeg'24	29	65.52%	55.17%	31.03%	27.59%
SBSeg'25	38	100.00%	84.21%	57.89%	71.05%

Table 1: Overview of badges awarded over the years.

[1] Dirk Beyer and Arnd Hartmanns. 2025. Reproducibility and replication of research results. *International Journal on Software Tools for Technology Transfer* (2025), 1–5.
[2] Dirk Beyer and Stefan Winter. 2025. Artifact Evaluations for Stronger Research Results. In *Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering*. 1234–1237.
[3] David Broneske, Sebastian Kittan, and Jacob Krüger. 2024. Sharing Software-Evolution Datasets: Practices, Challenges, and Recommendations. *Proceedings of the ACM on Software Engineering* 1, FSE (2024), 2051–2074.
[4] Association for Computing Machinery. 2025. Artifact Review and Badging Version 1.1. https://www.acm.org/publications/policies/artifact-review-and-badging-current. (2025). Accessed: 2025-10-31.
[5] Ben Hermann. 2022. What has artifact evaluation ever done for us? *IEEE Security & Privacy* 20, 5 (2022), 96–99.
[6] Ben Hermann, Stefan Winter, and Janet Siegmund. 2020. Community expectations for research artifacts and evaluation processes. In *Proceedings of the 28th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*. 469–480.
[7] Neil P Chue Hong. 2021. Reproducibility badging and definitions: A recommended practice of the national information standards organization. (2021).
[8] Daniel Olszewski, Allison Lu, Carson Stillman, Kevin Warren, Cole Kitroser, Alejandro Pascual, Divyajyoti Ukirde, Kevin Butler, and Patrick Traynor. 2023. "Get in Researchers; We're Measuring Reproducibility": A Reproducibility Study of Machine Learning Papers in Tier 1 Security Conferences. In *ACM SIGSAC conference on computer and communications security*. 3433–3459.
[9] Mohammad Reza Saleh Sedghpour, Alessandro Vittorio Papadopoulos, Cristian Klein, and Johan Tordsson. 2024. Artifact evaluation for distributed systems: Current practices and beyond. *arXiv preprint arXiv:2406.13045* (2024).