Artifacts Availability & Reproducibility
(VLDB 2021 Round Table)

Manos Athanassoulis
Boston University

Peter Triantafillou
University of Warwick

Raja Appuswamy
EURECOM

Rajesh Bordawekar
IBM Research

Badrish Chandramouli
Microsoft Research

Xuntao Cheng
Alibaba Cloud

Ioana Manolescu
Inria and Institut
Polytechnique de Paris

Yannis Papakonstantinou
Databricks and UCSD

Nesime Tatbul
Intel Labs and MIT

ABSTRACT

In the last few years, SIGMOD and VLDB have intensified efforts to encourage, facilitate, and establish reproducibility as a key process for accepted research papers, awarding them with the *Reproducibility* badge. In addition, complementary efforts have focused on increasing the sharing of accompanying artifacts of published work (code, scripts, data), independently of reproducibility, awarding them the *Artifacts Available* badge. In this short note, we summarize the discussion of a panel held during VLDB 2021 titled "Artifacts, Availability & Reproducibility". We first present a more detailed summary of the recent efforts. Then, we present the discussion and the contributed key points that were made, aiming to assess the reproducibility of data management research and to propose changes moving forward.

1. SUMMARY OF RECENT EFFORTS

Reproducibility. In the recent past, the goal of reproducing scientific results has been adopted by increasingly many research communities. By indicating that a published research paper is *reproducible* we refer to a result that obtained with stated precision by a different team using the same experimental setup and the artifacts provided by the author, following the current ACM terminology[1].

SIGMOD Reproducibility 2008-2012. In 2008, SIGMOD was the first data management conference that had an optional reproducibility[2] test of published papers [2]. In 2009, the effort was amended to test *Workability* as well, that is, whether the results can be generalized to broader experimental se-

tups [1][3]. This effort continued until 2012 where the goal was to make it part of the main conference and award to papers an official *Reproducible* badge for papers that have been independently reproduced and a *Sharable* badge for papers that have their artifacts available via a public URL[4].

SIGMOD Reproducibilty 2016-2021. This reproducibility effort stopped between 2012 and 2015, and was resumed in 2016[5]. The effort built on the previous instances of SIGMOD Reproducibility to create a standardized process for authors of accepted SIGMOD papers to follow. The *SIGMOD Reproducibilty Committee* (RC) was established (renewed periodically) and calls for papers to be submitted for reproducibility were issued. Each submission was assigned to a member of RC, which evaluated the extent to which presented results could be adequately reproduced by them, using the artifacts (scripts, code, and data) provided by authors. Since 2017, reproducibility awards have been awarded to the easiest to reproduce artifacts[6], which are presented during the conference event.

PVLDB Joins the Ranks in 2018. In September 2018, PVLDB launched its own reproducibility effort[7]. Similar to the SIGMOD effort, an RC was created and periodically renewed and authors of accepted papers were invited to undergo a reproducibility process, akin to that of SIGMOD. Also, awards were established and handed out to the authors of the easiest to reproduce papers.

---

[1] https://www.acm.org/publications/policies/artifact-review-and-badging-current

[2] Initially, the term *Repeatability* was used, however, since 2016, multiple communities converged to *Reproducibility*. Eventually, ACM adopted *Reproducibility* formally in 2020, which we use throughout this report to avoid confusion.

[3] The *Repeatability and Workability* evaluation continued for 2010 (https://event.cwi.nl/SIGMOD-RWE/2010/) and 2011 (https://event.cwi.nl/SIGMOD-RWE/2011/).

[4] https://www.sigmod.org/2012/reproducibility.shtml

[5] https://reproducibility.sigmod.org/history.html

[6] https://sigmod.org/sigmod-awards/sigmod-most-reproducible-paper-award/

[7] https://vldb.org/pvldb/reproducibility/

Currently, PVLDB mandates that, in addition to the voluntary nature of reproducibility for research papers, all experimental, analysis, and benchmark (EAB) papers *must* undergo reproducibility review.

Availability and Reproducibilty 2022. Starting from 2022, the SIGMOD effort is renamed to *Availability and Reproducibilty* and the committee renamed to *SIGMOD ARC*. Similarly, the VLDB RC is also conducting artifact availability checks.

Artifacts Availability as a Goal. Both VLDB and SIGMOD have been recognizing the increasing importance of artifact availability as a means to (1) increase the impact of database research papers, (2) enable easy dissemination and understanding of research results, and (3) enable easy sharing and uptake of data, code and experimentation set-ups. As a result, both conferences are strongly suggesting that all papers make their artifacts available, and the authors are asked to provide a justification when the do not submit links to their code and data. VLDB invites all papers to share a public URL as part of the paper metadata and the VLDB RC is carrying out checks for such links before awarding the badge. Starting from Volume 16, artifact availability for accepted papers in VLDB is considered part of the review process. SIGMOD invites all papers to share an archive with the artifacts which is also checked by SIGMOD ARC. Note that artifacts may include code, scripts, and experimental set-up on one hand, and also data sets, on the other hand.

Currently, the vetted artifacts along with deployment documentation, are made available through the ACM digital library as supplemental material. The data management community is also considering creating citable artifacts for code and data using independent DOIs. Similar efforts take place in sibling communities like systems, where conferences have an optional *call-for-artifacts* (e.g., SOSP[8], OSDI[9]), machine learning with their *reproducibility challenge*[10], information retrieval, where the flagship conference, SIGIR, has a track for reproducibility papers[11] and they assign badges and unique DOIs to code and data[12], and data mining, where reproducibility is part of the review process[13].

Participation. The participation of accepted pa-

---

[8] https://sysartifacts.github.io/sosp2021/call.html
[9] https://www.usenix.org/conference/osdi21/call-for-artifacts
[10] https://paperswithcode.com/rc2021
[11] https://sigir.org/sigir2022/call-for-reproducibility-track-papers/
[12] https://sigir.org/general-information/acm-sigir-artifact-badging/
[13] https://kdd.org/kdd2022/cfpResearch.html

pers in the reproducibility process has been fluctuating with mostly increasing trends over the last few years. For example, SIGMOD received 9 reproducibility submissions of 2019 papers, 16 for 2020, and 25 for 2021 (that are currently under reproducibility review). VLDB RC is on track to receive about 20 papers in this submission cycle. In fact, a key point brought up in the VLDB 2021 panel was how to increase participation. Since the effort to make artifacts available just started, we look forward to report on those in future reports and panels. But, there are encouraging signs: from the VLDB side, more than 120 papers published in Volume 14 carried the *Artifacts Available* badge. In the current issue of Volume 15, VLDB RC observed that about 70% of all published papers carried the *Artifacts Available* badge in their published pdf.

2. PANEL QUESTIONS

Is what we are doing at present science?
The panel had a lively discussion that highlighted that our community's research has a strong practical component, hence, availability of artifacts helps to increase the impact of published papers. The panel further discussed the importance for the community to test reproducibility in, ideally, all papers, as a clear signal that we value impact over quantity. Yannis Papakonstantinou highlighted that a large fraction of the data management papers are essentially technological contributions aiming to build new artifacts, hence artifact availability is key.

Should published results be reproduced?
Panelists agreed that reproducibility reviewing is desired for all research results, however, it was noted that there are artifacts that cannot be shared mostly for legal reasons (e.g., industrial closed-source systems and sensitive data sets). This can be mitigated if reproducibility uses the artifacts without publicizing them for these cases. For example, Xuntao Cheng pointed out that it should be an option to release the binary only and not to open source the code. Badrish Chandramouli highlighted that while availability and reproducibility are crucial, the community accepts claims at face value and should have a degree of trust in the authors.

Should we require availability?
The panel argued that availability and reproducibility are both crucial, and in fact, Rajesh Bordawekar, pointed out that code availability is absolutely necessary for showing impact on practical problems and accelerating research progress. Further, Rajesh proposed to elevate the artifacts to a separate *citable* entity with its own DBLP entry and potentially DOI, and hold sessions to highlight avail-

able and reproduced code artifacts. Badrish Chandramouli and Yannis Papakonstantinou pointed out that availability is not the same as *usability*, in the sense that publicly available artifacts should also have detailed instructions on how to use them to have impact. In fact, in order for a VLDB paper to receive the *Artifact Available* badge, the submission to the RC must be accompanied by an easy-to-read set of instructions describing the code and data sets and their use. Further, Badrish highlighted that industrial research might involve (modifications to existing) closed-source systems (e.g., the Microsoft SQL Server optimizer), which might be harder to be made available. He further pointed out that with some extra effort, such contributions could also validate the ideas using open-source alternatives, thereby releasing such artifacts. Currently, the authors of VLDB and SIGMOD submissions are asked to justify it when artifacts are not available.

Even then, is availability enough?

The panel agreed that availability is paramount and should be as wide as possible. Further, the panelists agreed that verifying reproducibility is also crucial for three reasons: (i) to reward well-developed, easy to use/adopt tools that help accelerate the progress of our field, (ii) send the signal that impact and applicability is valued by our community, and (iii) discourage poor scholarship and/or malicious publication efforts. However, both Raja Appuswamy and Xuntao Cheng pointed out that in addition to availability and reproducibility, we need cloud-based availability that can readily offer a library of prior artifacts in an easy-to-deploy setup. Badrish highlighted that one trend seen in industrial labs is to incorporate reproducibility in the development process, making it almost automatically reproducible. The main challenge is that sometimes new industrial research results are based on new proprietary hardware making it very hard to reproduce without access to that hardware.

Please comment on the status quo.

In this question, the panel discussed diverse points of view. Xuntao Cheng highlighted that publishing for industry is a means to advertise high-quality work and build technical reputation for the company. Further, he pointed out that our community needs to work on delivering more benchmarks that more closely match modern applications. Rajesh Bordawekar proposed to make the review process open (similar to NeurIPS) and double-blind and focus on code-reuse and availability. Raja Appuswamy pointed out that in other communities (e.g., Bioinformatics) the source code and the

data are *required* for the paper to be accepted, and the authors have to provide *usable* material in order for the paper to get published. Nesime Tatbul shared her view on making the process easier by giving incentives to everyone to participate (e.g., more visibility for the authors, recognition and education opportunities for reviewers, faster innovation for adopters, and sharing success stories from all fronts), and also work towards having a better time management for the authors through up-to-date best practices and shared infrastructure. This resembles the point raised earlier for a common research cloud with a library of approaches. Ioana Manolescu pointed out that currently the authors only perceive a small benefit from making sure their code is reproducible, while the effort to make it available and package it for reproducibility is frequently perceived as much higher. This resonates with Nesime's point about giving incentives for researchers (especially junior ones) to participate both as submitting authors and as reviewers.

Overall, the panelists agreed that reproducibility has made progress in our community, however, more steps still need to be taken.

How would you suggest to change the status quo?

The panelists came with a wealth of ideas on how to change the status quo, focusing mostly on making the process easier and smoother for the authors, the reviewers, and the (future) researchers that will be able to benefit from the available artifacts.

Incentivize. Virtually all the panelists agreed that it is important to give additional incentives to authors and reviewers to participate. The incentives can be both rewards and penalties. For example, Ioana supported that the extra recognition of the reproducibility and availability badges is not always enough. Her proposal is to ensure that at training and education level, all graduate students learn good coding practices early on and use the available state-of-the-art tools. In order to further reward artifacts, Ioana also proposed to give higher acceptance chances to papers that verify. Badrish agreed on providing incentives, but does not want to punish authors that do not share their code, which may be perfectly acceptable in many circumstances. Nesime proposed to further incentivize through more publicity of success stories and role models.

Use good practices. Building on Ioana's point above, there was a unanimous proposal to make sure that all authors should use good coding practices, code sharing tools (e.g., git), and if possible, build a repository of accepted solutions that researchers can quickly access artifacts (verified or not). The idea of a common repository (or collection of repos-

itories) got wide support in the discussion. In addition, Raja Appuswamy suggested that building a common *data management dev cloud* that will allow researchers to have access to virtual machines with high-end capabilities would significantly help this process. Further, tools like docker that help make execution and cloud deployment easy should also be employed. Yannis Papakonstantinou further highlighted that it is very important to ensure that artifacts are always accompanied by the setup, the parameters, and all the necessary details to use and reproduce an artifact.

Learn from other communities. Several of the panelists highlighted that there is expertise in other communities that can be leveraged. For example, Nesime and Rajesh mentioned the NeurIPS efforts in reproducibility, highlighting the open access journal *ReScience*[14] that publishes reproducibility reports, and the *reproducibility challenge* as a tool to crowd-source the effort and educate young researchers. Similarly, the systems community has integrated artifact evaluation in all of its conferences as an optional but strongly suggested step.

Increase experimental analysis papers. Yannis Papakonstantinou highlights that from the papers published in our community only a small fraction is *evaluation* papers (in the form of experimental analysis and benchmarking papers), while in other disciplines (e.g., Medicine) it is frequent that the *evaluation* papers are a much larger fraction of all published papers. One actionable item is to welcome more such evaluation papers, as VLDB has pioneered for a few years now.

A verifiability metric. Lastly, Yannis proposed the addition of a new review metric called *verifiability* to capture the reviewer's confidence that the available details and artifacts are enough to reproduce and verify the presented results.

3. CONCLUSION & FUTURE PLANS

The efforts surrounding reproducibility and artifact availability are now maturing steadily within our community, as evidenced by the SIGMOD and VLDB efforts. Artifact availability and reproducibility are increasingly being sought after, which is shown by the increasing participation in the SIGMOD and VLDB reproducibility processes. Regarding artifact availability, both SIGMOD and VLDB publish artifacts as supplemental material, while creating dedicated entries for software and code with unique DOIs is also considered. This approach would allow artifacts to be directly citable. However, storing a snapshot of the artifacts (the one

associated with the published paper) as a separate entity makes it harder to cite new versions of the artifacts. Further, by having two citable entities, the research paper that explains intricacies and details of the artifact might be less accessed or cited. A plan under consideration moving forward is to continue to attach software artifacts to the paper, and create unique DOIs for artifacts that contain new data sets. Ultimately, our goal is for all published papers to have their code, scripts and data available, and ideally to have been reproduced.

However, there are obstacles along the way; the primary one being the significant additional workload imposed on authors and reviewers, especially to carry out reproducibility checks. The panel concluded that more effort on rewarding this workload is needed to ensure that the reproducibility process will further expand and eventually become an integral part of the publication process. We are far from the ideal scenario where every single published paper in our community has successfully undergone a reproducibility process and is accompanied by artifacts (when possible). Similar to the recommendations of the National Academies report[15] [3], to reach our ideal state, we need (i) to inform researchers about the need and associated benefits, (ii) to blend artifact generation as a first-class citizen within the education and training of our future colleagues, and (iii) create the necessary operational infrastructures. Overall, our efforts are part of a larger effort at a higher level, exemplified by the new ACM *Reproducibility Emerging Interest Group*[16] that hosts an annual workshop[17] dedicated on practical reproducibility. The data management research community is determined to continue and deepen these efforts.

For any questions and suggestions, please contact the SIGMOD ARC and VLDB RC chairs.

4. REFERENCES

[1] S. Manegold, I. Manolescu, L. Afanasiev, J. Feng, G. Gou, M. Hadjieleftheriou, S. Harizopoulos, P. Kalnis, K. Karanasos, D. Laurent, M. Lupu, N. Onose, C. Ré, V. Sans, P. Senellart, T. Wu, and D. E. Shasha. Repeatability & workability evaluation of SIGMOD 2009. *SIGMOD Rec.*, 38(3):40–43, 2009.
[2] I. Manolescu, L. Afanasiev, A. Arion, J. Dittrich, S. Manegold, N. Polyzotis, K. Schnaitter, P. Senellart, S. Zoupanos, and D. E. Shasha. The repeatability experiment of SIGMOD 2008. *SIGMOD Rec.*, 37(1):39–45, 2008.
[3] National Academies of Sciences Engineering and Medicine. *Reproducibility and Replicability in Science.* The National Academies Press, Washington, DC, 2019.

---

[14] http://rescience.github.io/

[15] https://www.nationalacademies.org/our-work/reproducibility-and-replicability-in-science
[16] https://reproducibility.acm.org/
[17] https://p-recs.github.io/2022/