

Let's Stop Building at the Feet of Giants: Recovering unavailable Requirements Quality Artifacts

Julian Frattini¹, Lloyd Montgomery², Davide Fucci¹, Jannik Fischbach^{3,4},
Michael Unterkalmsteiner¹ and Daniel Mendez^{1,4}

¹Blekinge Institute of Technology, Valhallavägen 1, 371 41 Karlskrona, Sweden

²University of Hamburg, 20146 Hamburg, Germany

³Netlight Consulting GmbH, Sternstraße 5, 80538 Munich, Germany

⁴fortiss GmbH, Guerickestraße 25, 80805 Munich, Germany

Abstract

Requirements quality literature abounds with publications presenting artifacts, such as data sets and tools. However, recent systematic studies show that more than 80% of these artifacts have become unavailable or were never made public, limiting reproducibility and reusability. In this work, we report on an attempt to recover those artifacts. To that end, we requested corresponding authors of unavailable artifacts to recover and disclose them according to open science principles. Our results, based on 19 answers from 35 authors (54% response rate), include an assessment of the availability of requirements quality artifacts and a breakdown of authors' reasons for their continued unavailability. Overall, we improved the availability of seven data sets and seven implementations.

Keywords

requirements quality, open science, availability, artifacts, data set

1. Introduction

Data sets and tools are often reported as important contributions to requirements quality literature [1]. However, a recent secondary study revealed that out of 57 primary studies, as little as 12% of data sets and 19% of tools are currently publicly available [2]. The unavailability of those artifacts has two major consequences. Firstly, empirical results are difficult to *reproduce*, which inhibits the process of strengthening the empirical evidence of scientific contributions. Secondly, the presented artifacts are difficult to *reuse*, which necessitates scientific progress to

In: A. Ferrari, B. Penzenstadler, I. Hadar, S. Oyedeji, S. Abualhaija, A. Vogelsang, G. Deshpande, A. Rachmann, J. Gulden, A. Wohlgemuth, A. Hess, S. Fricker, R. Guizzardi, J. Horkoff, A. Perini, A. Susi, O. Karras, A. Moreira, F. Dalpiaz, P. Spoletini, D. Amyot. *Joint Proceedings of REFSQ-2023 Workshops, Doctoral Symposium, Posters & Tools Track, and Journal Early Feedback Track. Co-located with REFSQ 2023. Barcelona, Catalunya, Spain, April 17, 2023.*

✉ julian.frattini@bth.se (J. Frattini); lloyd.montgomery@uni-hamburg.de (L. Montgomery); davide.fucci@bth.se (D. Fucci); jannik.fischbach@netlight.com (J. Fischbach); michael.unterkalmsteiner@bth.se (M. Unterkalmsteiner); daniel.mendez@bth.se (D. Mendez)

🌐 <https://lloydm.io/> (L. Montgomery); <https://dfucci.github.io/> (D. Fucci); <https://www.lmsteiner.com/> (M. Unterkalmsteiner); <https://www.mendezfe.org/> (D. Mendez)

🆔 0000-0003-3995-6125 (J. Frattini); 0000-0002-8249-1418 (L. Montgomery); 0000-0002-0679-4361 (D. Fucci); 0000-0002-4361-6118 (J. Fischbach); 0000-0003-4118-0952 (M. Unterkalmsteiner); 0000-0003-0619-6027 (D. Mendez)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

restart over and over again instead of evolving from existing contributions. Ultimately, these consequences inhibit the progress of the requirements quality research domain.

Following open science practices in software engineering improves the accessibility of artifacts and the preservation of future contributions [3] but it is difficult to apply them in retrospect. Because of this, the accessibility of artifacts in past publications deteriorated over time [1]. The resulting unavailability of artifacts [2] poses a significant challenge to artifact-dependent research, like the requirements quality or the larger natural language processing for requirements engineering (NLP4RE) domain. In this work, we set to recover unavailable requirements quality artifacts by requesting authors to disclose them according to open science principles. Our main contribution is the recovery of seven data sets and seven implementations. Nevertheless, 16 out of 35 (46%) requests to authors remained unanswered, indicating that the research community needs to emphasize further the importance of persistently archiving research data.

The remainder of this manuscript is structured as follows: Section 2 introduces the background both on the topic of open science and requirements quality research. Section 3 describes the process and Section 4 the results of the recovery. We discuss these results in Section 5 before concluding in Section 6.

2. Background

2.1. Open Science in Software Engineering

Scientific work needs to be reproducible [3] to strengthen the evidence it contributes to a field of research [4]. *Open science* is the initiative of ensuring public availability of research artifacts [3] and, hence, facilitating reproducibility. Within open science, the facets of *open access* for publications, *open data* for data sets, and *open source* for source code are most relevant to software engineering [5], where each facet of open science entails different techniques and best practices to disclose its respective type of research artifact. Several governmental research funding agencies, including the European Union, made open access to scientific results (including data, tools, etc.) mandatory¹.

In literature, Minocher et al. propose four attributes *data recoverability*, *data usability*, *analytical clarity*, and *agreement of results* and explicitly emphasize the sequential dependency of those attributes—e.g., *analytical clarity* of data is meaningless if the data is not *recoverable*.

Recent endeavors of incentivizing scholars to follow open science principles include open science badges [6] and the registered reports [7]. However, the software engineering research community is still in the process of adapting open science principles [3], and the unavailability of artifacts is common [1]. Prominent reasons for the unavailability of artifacts include the sensitivity of data or corresponding authors changing their affiliation and consequently losing access to their artifacts [8]. While some reasons for the unavailability of artifacts (e.g., the sensitivity of company-owned data) may well require significant effort to cope with, other reasons (e.g., loss of artifact, lack of diligence) can be circumvented easily by following proposed guidelines [3] and making use of modern tools for artifact sharing.

¹https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-access_en

2.2. Requirements Quality Literature

Recent advances have established that artifacts produced in requirements engineering (RE) have a significant impact on downstream software development activities [9], potentially even causing project failure [10]. Consequently, requirements artifacts merit quality assurance [1]. The requirements quality literature is dedicated to providing the understanding as well as the support for measuring and improving the quality of requirements [1]. One popular approach to this is the proposal of *quality factors*. Requirements quality publications often formulate one or more quality factors—e.g., the use of *coordination ambiguity* leading to divergent interpretations [11]—annotate instances of that quality factor in a *data set*, and finally present an *implementation* (i.e., an algorithm or full-fledged tool) to detect these instances automatically.

These artifacts—both data sets and implementations—represent essential contributions facilitating empirical research and technology transfer. While the (annotated) data sets are the main driver for developing new and improving existing implementations for quality factor detection, implementations are the tools to be deployed in industry for actual integration and improvement of the software engineering process. The NLP4RE research domain, which applies natural language processing (NLP) techniques to RE [12] and constitutes a large part of the contributions to the requirements quality literature [1], is particularly focused on said delivery and improvement of tools. In addition to the dependency of these NLP-powered tools on the availability and reliability of training data, this puts the NLP4RE research domain on the forefront of the open science challenge [12]. The NLP4RE community is therefore particularly aware of its dependency on the availability of artifacts [13].

However, recent systematic studies revealed that a significant amount of these artifacts are not available² anymore or have never been [12, 1, 2]. Table 1 reports the availability status of 57 data sets (D) and 36 implementations (I) extracted from the 57 primary studies of our previously-published literature review on requirements quality factors [2].

3. Recovery Process

The insight that the availability of requirements quality artifacts is insufficient [1, 2] motivated our objective to improve the state of open science in the requirements quality literature by ensuring the *recoverability of data*, a necessary prerequisite for the reproducibility of scientific work [14].

In this section, we document the artifact recovery process along the undertaken steps. In Section 3.1, we describe the selection of the sample of primary studies. We detail our approach to contact corresponding authors in Section 3.2 and maintain correspondence with them in Section 3.3. Finally, we document the evaluation of the recovery process and success in Section 3.4. All produced data, scripts, and documentation are disclosed in our replication package⁴.

²Where available means a status of *Upon request* (see Table 1) or better.

³In this context, we are using the term *open source* as commonly understood, not as used in the open science framework [5], which would imply adherence to the properties listed under *open data*.

⁴Available at <https://doi.org/10.5281/zenodo.7708571>

Table 1

Availability status of requirements quality artifacts [2] including data sets (D) and implementations (I)

Status	Explanation	(D)	(I)
Open Data	The artifact is hosted in a service that satisfies the following criteria: (1) immutable URL (cannot be altered by the author or someone else), (2) permanent (the hosting organization has a mission to maintain artifacts for the foreseeable future), (3) accessible (there is a DOI pointing to the real data source URL), and (4) open-source license (the artifact has a license which grants access and re-use)	1	0
Open Source ³	[only for implementations] The implementation is available for all to use, and the code base has been disclosed	-	5
Available in paper	[only for data sets] The data set is small enough that the authors disclose the entire data set in the manuscript	5	-
Reachable link	The artifact is reachable now but is missing some of the Open Data aspects (see above)	1	1
Upon request	Authors claim the artifact is available upon request	0	1
Broken link	A link to the artifact is contained in the paper, but it does not resolve	10	1
No link	An artifact is presented, but no indication on how to access it is provided	15	27
Private	The authors state that an artifact exists but is private for some reasons (such as industry collaboration with private data, etc.)	24	0
Proprietary	The artifact is proprietary, and access is granted upon payment	1	1
Total		57	36

3.1. Study sample selection and preparation

We used *convenience sampling* since the primary studies on which we base our results were selected based on expediency [15]. In particular, we recovered artifacts from a set of primary studies used to build an ontology of requirements quality factors [2]. To develop this ontology, we collected manuscripts reporting *quality factors* from an original set of publications reported in another secondary study [1]. Extracting data sets and implementations from such publications revealed the unfortunate state of artifact availability.

We enhanced the data regarding data sets and implementations from our previous study [2] with the following information.

- Corresponding author: each artifact was associated with a corresponding author.
- Mention: each artifact was associated with its verbatim mention in the manuscript.

Additionally, we corrected information about one data set and three implementations that persisted in the previous study [2].

Using a spreadsheet, we collected data about

1. **authors** (n=35), specifying for each author the name and email address,
2. **data sets** (n=57), specifying for each data set its containing publication, its verbatim mention, the corresponding author, and its current availability, and
3. **implementations** (n=36), specifying for each implementation its containing publication, its verbatim mention, the corresponding author, and its current availability.

3.2. Approaching authors

We created a Python script that automatically assembles one email for each corresponding author. This email contained the following elements:

1. Header: an explanation of our endeavor and a request to contribute to open science (or alternatively explain why this is impossible).
2. Artifact list: a list of artifacts contained in the publications of the authors that were not open access.
3. Instructions: brief *how to* to properly disclose artifacts according to the open science principles as well as the offer to assist them in the process
4. Contact: a way to reach out to us.

We approached the authors in a first mail on the 30th of November 2022, followed by a reminder on the 13th of December, and a final reminder on the 11th of January 2023. For authors that did not respond to our request until the final reminder, we additionally contacted their co-authors to increase the likelihood of response. We concluded the recovery process on the 8th of February 2023, yielding a time frame of 70 days.

3.3. Correspondence

We kept close contact with the authors we approached by responding in a window of 24 hours within workdays. During this process, we clarified concerns and offered our help. We processed and recorded the information contained in the authors' answers in a spreadsheet file. We tracked the *response status* in an additional column, denoting the request as either *undeliverable*, *unanswered*, *answered*, or *completed*. We labeled a recovery request as completed once the corresponding author, for all their artifacts, either improved their availability or explained the inability to recover or disclose them.

Furthermore, we documented the dates of the first email sent, the first response received, and the completion of the request alongside the number of emails sent by the author in addition to the updated availability status of the artifacts and, eventually, the author's explanation for not taking the recommended actions. Two authors coded these explanations independently and came to an absolute agreement on the types of reasons for non-recovery. When the corresponding author's email address was no longer used, we reached out via personal contacts or social networks like Twitter and LinkedIn.

3.4. Evaluation

To evaluate the artifact recovery process, we generated statistics of the following data from the documentation in our tables.

1. Correspondence (author response time and frequency) to evaluate the effort of the recovery process.
2. Recovery request success (change in artifact availability) to evaluate the success of the recovery process.

- Reason for non-recovery (author responses excusing the recovery) to evaluate the reasons inhibiting open access.

We evaluated the data by generating descriptive statistics from our documentation.

4. Results

4.1. Correspondence

Out of the 35 approached corresponding authors, 19 answered the recovery request, and 13 completed it. We could not reach three authors despite searching for a valid contact. The distribution of correspondence status is visualized in Figure 1a. It took, on average, 14.6 days for a corresponding author to reply to our request and 22.4 additional days to complete the request. On average, a request was resolved in an exchange of 3 emails with the corresponding author. The distributions of these statistics are visualized in Figure 1b and Figure 1c, respectively.

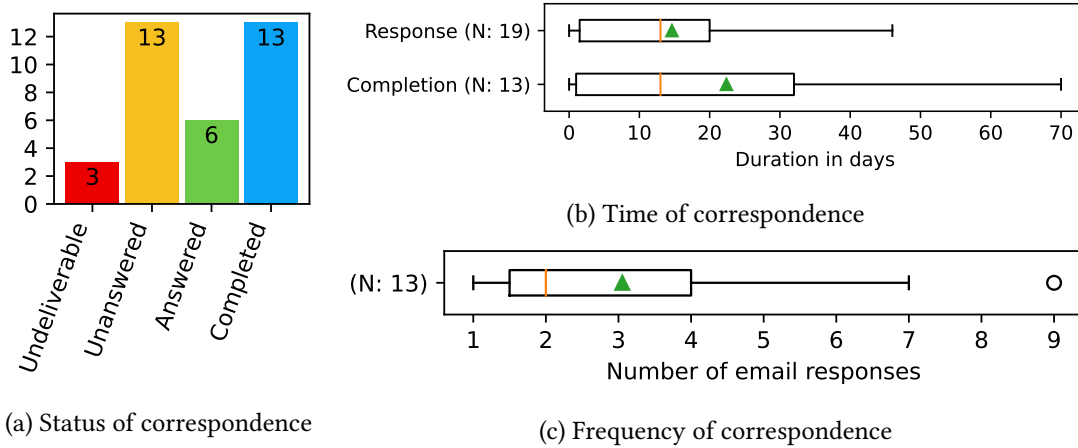


Figure 1: Correspondence to the artifact recovery request

4.2. Artifact Recovery Success

The corresponding authors improved the availability of seven data sets (four of which follow open-access principles) and seven implementations (six following open-access principles). This increases the availability of data sets from 12.3% (7/57, 1 open access) to 22.8% (13/57, 5 open access) and the availability of implementations from 19.4% (7/36, 0 open access) to 30.6% (11/36, 6 open access). Authors further confirmed the unavailability of 21 data sets and six implementations and provided reasons for the inability to recover or disclose them.

Figure 2 visualizes the success of the recovery request. The heatmap considers all artifacts (data sets in Figure 2a and implementations in Figure 2b) where the corresponding author completed the recovery request. The number in a cell represents the number of artifacts for which the original availability (on the y-axis) has been updated to the new availability (on the x-axis). The count of artifacts whose availability remained the same (e.g., because an author

confirmed that the artifact could not be made more available) is reported on the diagonal (shaded gray). An improvement in the availability of an artifact contributes to cells to the right of the diagonal, a deterioration of the availability to the left.

For example, one implementation was previously available *upon request* [16]. Now that the authors disclosed the implementation following open access principles⁵, the entry moved three cells to the right (see Figure 2b).

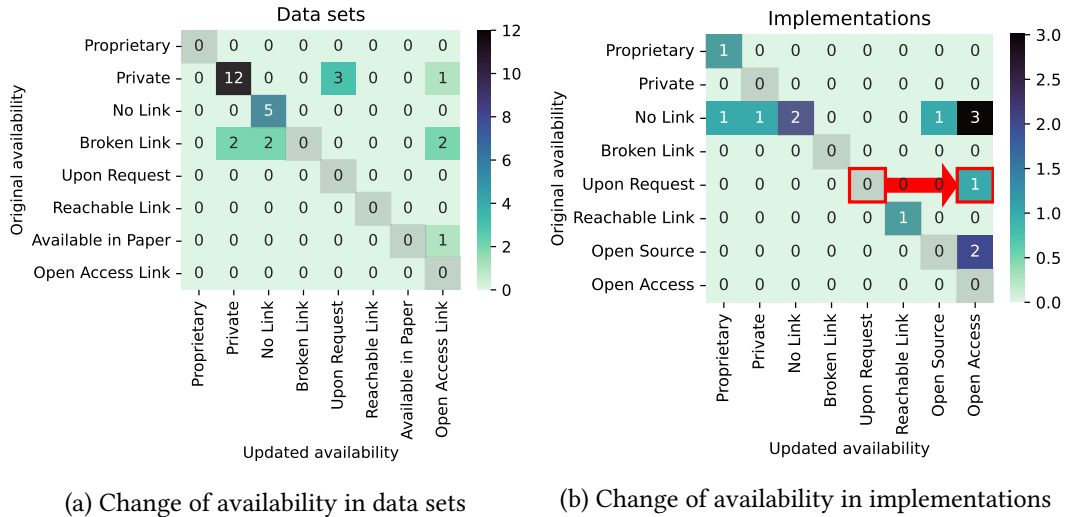


Figure 2: Change of artifact availability

The inability to recover or disclose artifacts was reported as follows: among 21 unrecoverable data sets, 15 were lost (i.e., the author could not find them anymore or the contact, whom the author assumed had the data, was unreachable), and six could not be disclosed due to sensitive contents. Among the six unrecoverable implementations, three became proprietary, and three were lost.

5. Discussion

Within the 70-day time frame for the process, authors of requirements quality publications recovered several data sets and implementations that are now available for reproduction of scientific results and reuse in future projects. We referenced the recovered artifacts in the requirements quality factor ontology [2]⁶ as well as our replication package to make them accessible.

Additionally, the authors confirmed the unavailability of several more artifacts. While this does not actively improve the availability of artifacts for reproduction, it clarifies the ambiguous status of several data sets and implementations. Overall, when authors answered a recovery request, they either recovered their data or reported the inability to do so with helpful

⁵Now publicly available at <https://doi.org/10.5281/zenodo.7484023>

⁶See *Content* at <http://reqfactoront.com/>

explanations. Recovery requests failed due to (1) no response, (2) the artifact being lost, or (3) the artifact containing sensitive information. We did not encounter other reasons for the failure of a recovery request, which corroborates the goodwill of the sampled requirements quality community in its commitment to open science. This stands in contrast to the experience of other artifact recovery attempts, where researchers encountered reasons like requests for reimbursements or not seeing any personal gain in the recovery [8].

We cannot claim that our observations are universally valid for the software (requirements) engineering community due to the limitations of our study. For one, the set of primary studies was obtained via convenience sampling from a previous study [2]. This sample has known limitations as several primary studies relevant to requirements quality literature are missing. Hence, the results of recovery success and correspondence do not represent the complete requirements quality literature and research community. Furthermore, our conclusion regarding the status of correspondence, especially the status of *unanswered* and *answered* requests, is limited by how we decided to approach corresponding authors. Using emails as the mean of communication impedes the response rate since they are often abandoned with a change of affiliation [17]. The limited success of correspondence is a consequence of the time frame and communication channels used in this study rather than an indicator of the research community's attitude towards open science.

6. Conclusion

Both the credibility and reusability of previous publications in the requirements quality literature have been impeded by the unavailability of data sets and implementations. We requested corresponding authors of 57 publications to disclose their artifacts according to the open science principles. We improved the availability of seven data sets and seven implementations, several of which now follow open science principles.

With this study, we want to raise awareness about the importance of recovering artifacts associated with older publications. While adherence to the open science principles recently rose thanks to comprehensive guidelines (see [3]) or community initiatives such as artifact evaluation tracks at conferences, they are rarely applied retroactively to previous publications. Furthermore, we hope that the material we created will support researchers in areas that heavily rely on artifacts, such as NLP4RE, to recover more of them.

Our agenda, in the scope of the requirements quality factor ontology⁷, includes providing a central repository of updated information on the availability and location of relevant artifacts. We invite researchers to contribute to this cause and strengthen the evidence in our field.

Acknowledgments

The KKS foundation supported this work through the S.E.R.T. Research Profile project at Blekinge Institute of Technology. We additionally thank the reviewers for their valuable feedback upon which the manuscript was improved.

⁷<http://reqfactoront.com>

References

- [1] L. Montgomery, D. Fucci, A. Bouraffa, L. Scholz, W. Maalej, Empirical research on requirements quality: a systematic mapping study, *Requirements Engineering* (2022) 1–27.
- [2] J. Frattini, L. Montgomery, J. Fischbach, M. Unterkalmsteiner, D. Mendez, D. Fucci, A live extensible ontology of quality factors for textual requirements, in: *2022 IEEE 30th International Requirements Engineering Conference (RE)*, IEEE, 2022, pp. 274–280.
- [3] D. Mendez, D. Graziotin, S. Wagner, H. Seibold, Open science in software engineering, in: *Contemporary empirical methods in software engineering*, Springer, 2020, pp. 477–501.
- [4] B. C. Anda, D. I. Sjøberg, A. Mockus, Variability and reproducibility in software engineering: A study of four companies that developed the same system, *TSE* 35 (2008) 407–429.
- [5] J. Tennant, J. Beamer, J. Bosman, B. Brembs, N. C. Chung, G. Clement, T. Crick, J. Dugan, A. Dunning, et al., *Foundations for open scholarship strategy development* (2019).
- [6] M. C. Kidwell, L. B. Lazarević, E. Baranski, T. E. Hardwicke, S. Piechowski, L.-S. Falkenberg, C. Kennett, A. Slowik, et al., Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency, *PLoS biology* 14 (2016) e1002456.
- [7] B. A. Nosek, C. R. Ebersole, A. C. DeHaven, D. T. Mellor, The preregistration revolution, *Proceedings of the National Academy of Sciences* 115 (2018) 2600–2606.
- [8] M. Gabelica, R. Bojčić, L. Puljak, Many researchers were not compliant with their published data sharing statement: mixed-methods study, *Journal of Clinical Epidemiology* (2022).
- [9] S. Wagner, D. M. Fernández, M. Felderer, A. Vetrò, M. Kalinowski, R. Wieringa, D. Pfahl, T. Conte, M.-T. Christiansson, D. Greer, et al., Status quo in requirements engineering: A theory and a global family of surveys, *TOSEM* 28 (2019) 1–48.
- [10] D. Mendez, S. Wagner, Naming the pain in requirements engineering: Design of a global family of surveys and first results from germany, in: *Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering*, 2013, pp. 183–194.
- [11] S. Ezzini, S. Abualhaija, C. Arora, M. Sabetzadeh, L. C. Briand, Using domain-specific corpora for improved handling of ambiguity in requirements, in: *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, IEEE, 2021, pp. 1485–1497.
- [12] L. Zhao, W. Alhoshan, A. Ferrari, K. J. Letsholo, M. A. Ajagbe, E.-V. Chioasca, R. T. Batista-Navarro, Natural language processing for requirements engineering: A systematic mapping study, *ACM Computing Surveys (CSUR)* 54 (2021) 1–41.
- [13] F. Dalpiaz, A. Ferrari, X. Franch, C. Palomares, Natural language processing for requirements engineering: The best is yet to come, *IEEE software* 35 (2018) 115–119.
- [14] R. Minocher, S. Atmaca, C. Bavero, R. McElreath, B. Beheim, Reproducibility improves exponentially over 63 years of social learning research (2020).
- [15] S. Baltés, P. Ralph, Sampling in software engineering research: A critical review and guidelines, *Empirical Software Engineering* 27 (2022) 1–31.
- [16] F.-L. Li, J. Horkoff, L. Liu, A. Borgida, G. Guizzardi, J. Mylopoulos, Engineering requirements with desiree: An empirical evaluation, in: *International Conference on Advanced Information Systems Engineering*, Springer, 2016, pp. 221–238.
- [17] J. D. Wren, J. E. Grissom, T. Conway, E-mail decay rates among corresponding authors in medline: The ability to communicate with and request materials from authors is being eroded by the expiration of e-mail addresses, *EMBO reports* 7 (2006) 122–127.