

Quality Guidelines for Research Artifacts in Model-Driven Engineering



Carlos Diego Nascimento Damasceno and Daniel Strüber

Institute for Computing and Information Sciences

Radboud University, Nijmegen, Netherlands

{d.damasceno,d.strueber}@cs.ru.nl

Abstract—Sharing research artifacts is known to help people to build upon existing knowledge, adopt novel contributions in practice, and increase the chances of papers receiving attention. In Model-Driven Engineering (MDE), openly providing research artifacts plays a key role, even more so as the community targets a broader use of AI techniques, which can only become feasible if large open datasets and confidence measures for their quality are available. However, the current lack of common discipline-specific guidelines for research data sharing opens the opportunity for misunderstandings about the true potential of research artifacts and subjective expectations regarding artifact quality. To address this issue, we introduce a set of guidelines for artifact sharing specifically tailored to MDE research. To design this guidelines set, we systematically analyzed general-purpose artifact sharing practices of major computer science venues and tailored them to the MDE domain. Subsequently, we conducted an online survey with 90 researchers and practitioners with expertise in MDE. We investigated our participants' experiences in developing and sharing artifacts in MDE research and the challenges encountered while doing so. We then asked them to prioritize each of our guidelines as essential, desirable, or unnecessary. Finally, we asked them to evaluate our guidelines with respect to clarity, completeness, and relevance. In each of these dimensions, our guidelines were assessed positively by more than 92% of the participants. To foster the reproducibility and reusability of our results, we make the full set of generated artifacts available in an open repository at <https://mdeartifacts.github.io/>.

Index Terms—Software artifacts, Open Science, Model-Driven Engineering, Quality Management

I. INTRODUCTION

The term “reproducibility crisis” has gained traction as researchers from various fields have reported challenges to reproduce studies [1]. Software engineering (SE) research is by no means exempt from this phenomenon, as many authors in SE have faced difficulties when trying to reproduce studies [2], performing replications with the direct support of authors [3]; and reusing artifacts in their own benchmarks [4]. To mitigate these issues, SE conferences have started to incorporate artifact evaluation processes [5]–[9]. In parallel to this, the Association for Computing Machinery has recently launched the *ACM SIGSOFT Empirical Standards* document to communicate frequent expectations for research methods commonly used by their community [10].

Software artifacts are known to provide means to others to build upon existing knowledge [11], adopt novel contributions in practice [12], and increase the chances of papers receiving attention [13]. However, the current lack of discipline-specific

guidelines for research data management [14] opens the opportunity for conflicting subjective expectations toward artifact quality and hence, misunderstandings on the true potential of research artifacts [15].

In Model-Driven Engineering (MDE), openly providing research artifacts plays a key role for the following reasons. First, despite some attempts to provide sets of models in consolidated repositories [16], [17] and collections of UML models [18], [19] and transformations [20], [21], there is still a lack of large datasets of models of diverse modeling languages and application domains [22]. Having more systematic artifact sharing practices would help to increase the potential reuse of available models and support the evaluation of research tools and techniques. Second, the need for consolidated artifact sharing practices in MDE research has recently become more pronounced, as the community targets a broader use of artificial intelligence (AI) techniques. Thus, to benefit from the advances in machine learning and even more so deep learning, MDE researchers should address the need for large open datasets and confidence measures for their quality.

In this paper, we present a set of guidelines for artifact sharing specifically tailored to MDE research. We designed and applied a comprehensive methodology to inform the design of our guidelines. This methodology included an analysis of available discipline-independent practices for artifact sharing from various SE venues, a study of literature on artifacts in the MDE domain, an application of project management principles from the literature, and an online survey with 90 practitioners and researchers with experience in MDE. We address the following research questions:

- RQ1:** How can one define domain-specific guidelines for artifact sharing in the MDE domain?
- RQ2:** How do the defined guidelines address the main challenges encountered by MDE experts?
- RQ3:** How do MDE experts prioritize our practices?
- RQ4:** What is the quality of the proposed guidelines?

The outcome is a set of 84 practices, structured along seven factual questions (i.e., what, why, where, who, when, how, how much/many), and prioritized into the three levels *essential*, *desirable*, and *unnecessary*. The quality of our guidelines in terms of completeness, clarity, and relevance was assessed positively by more than 92% of our participants. These findings indicate that our guidelines can be useful for guiding authors

in preparing high-quality research artifacts. Furthermore, we believe that our guidelines can inform organizers of artifact evaluation processes of MDE conferences and journals on questions to keep in mind when reviewing artifacts and on practices generally accepted in the community.

This paper is organized as follows: In Sect. II, we discuss the background concepts that underpinned our study. In Sect. III, we present the methodology used to address our investigations. In Sect. IV, we report the results of our survey and participants' responses. In Sect. V, we discuss the implications of our findings to artifact authors and reviewers, opportunities for improvement, and threats to validity. In Sect. VI, we enumerate some related work. In Sect. VII, we close this paper and discuss future work.

II. BACKGROUND

A. Research Data Management and Artifact Sharing

As Computer Science (CS) becomes more important in all branches of the modern sciences, software artifacts are now seen as another important outcome of scientific research [15]. They are necessary evidence to validate findings and results in most of modern sciences [23], help researchers building upon existing knowledge [11], encourage the adoption of novel research in practice [12], and increase the chances of citations [13]. However, As numerous SE researchers report increasing challenges to reproduce studies [2]–[4], the “reproducibility crisis” known in natural sciences and humanities [1] is now an imminent threat to the reproducibility, replicability, and repeatability of empirical experimentations in SE [24].

Research Data Management (RDM) helps investigators to make conscious decisions about research data [25]. It comprises various activities, from data planning and synchronization, over elements standardization, to data sharing and repository management; which encourage research reproducibility and software sustainability [26]. Funding agencies and research councils have started to incorporate requirements for data management and sharing as a standard part of their policies. Such initiatives can benefit research environments to promote their scholarly work, enable new collaborations, allow maximum use of contributed information, and advance science to the benefit of the whole society [27]. However, due to the multi-faceted nature of software artifacts in terms of purpose, size, complexity and format, RDM policies may still need to be adapted [14] to provide clear, complete and relevant domain-specific guidance.

B. Artifacts in Model-Driven Engineering

In Model-Driven Engineering, rigorous abstractions models are used as central artifacts in systems' design and development process [28]. Typical artifacts in MDE include models, metamodels, model transformations, and modeling tools [29]. As such distinct assets, MDE artifacts should comply with their requirements and quality criteria, such as model semantics and syntax [30] and technology standards [31]. Basso et al. [29] suggest that MDE artifacts should report information about aspects such as their context of use, usage restrictions,

and business opportunities. On top of technical information, organizational and social factors should also be considered as sources of issues that may affect the adoption of MDE tools and artifacts [32], such as sustainability over the long term, appropriateness for repurposing, and opportunities of interacting with authors/community. Studies like these constitute useful resources to guide MDE researchers and practitioners on creating and sharing high-quality research datasets, even more so as the MDE community aims at developing effective and efficient AI-based MDE techniques, a task that can only become feasible if large, open, and good-enough benchmark sets are available.

C. Project Quality Management

In a survey of more than 1.5k PhD students from Europe and Israel, the majority of the doctoral students reported having no training or expertise in managing research projects [33]. As doctoral research projects meet the same formal definition of a project [34], researchers suggest that PhD students should receive basic project management lessons to better manage their research as they move towards the successful completion of their doctorate [33].

According to the Project Management Institute (PMI), project management (PM) is defined as the application of knowledge, skills, tools, and techniques to project activities to meet the project requirements [34]. Project quality management is a PM knowledge area that applies to all projects, regardless of the nature of their deliverables (i.e., artifacts). It aims at incorporating the organization's quality policy regarding planning, managing, and controlling project and product quality requirements so that stakeholders' expectations are met. Quality measures and techniques are specific to the type of artifact being produced and should be always identified and documented *a priori*.

Plan quality management is the process of identifying quality requirements and/or standards for a project and its artifacts and documenting means that a project shall demonstrate compliance with such quality requirements and/or standards. Failure to identify and meet quality requirements can have serious negative consequences for project stakeholders. To understand the concept of quality within the context of a specific project, there is an extensive toolbox of methods [35] from which the 5W2H method constitutes a simple but powerful framework for research planning, analysis, or reviewing.

The term “5W2H” is an abbreviation of seven keywords: **What**, **Where**, **Why**, **Who**, **When**, **How**, and **How Much**. This is a well-known method by journalists to report news [36], [37]. From their perspective, reporters are expected to gather and present categories of information to their audience. These categories should indicate the essential information that people may want to know about a news story. In the literature, authors also refer to this method as the Five Ws [37] or 5W1H [36].

In project quality management [34], the 5W2H method can be used for asking questions about a process or problem. The 5W2H structure forces managers to consider all aspects

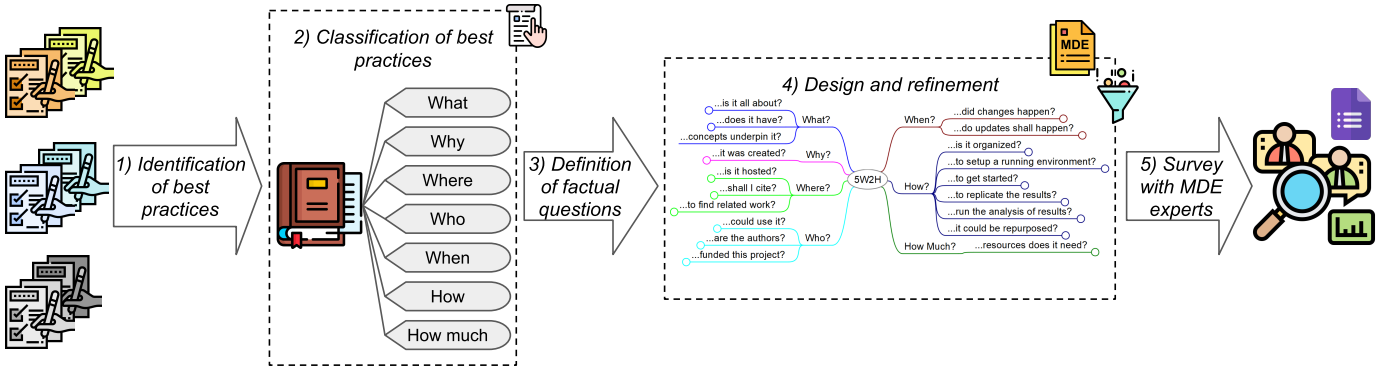


Fig. 1: Designing a set of quality guidelines for MDE research artifact sharing

of the situation when analyzing a process for improvement opportunities, a problem is identified but must be better defined, planning a project or steps of a project, or reviewing a project after completion [35].

According to Tague’s book “The Quality Toolbox” [35], the 5W2H method works as follows: (i.) Review the situation under study and make sure the subject of the 5W2H is understood. (ii.) Develop appropriate factual questions about the situation for each keyword. (iii.) Answer each question. If answers are not known, create a plan for finding them. (iv.) What you do next depends on your situation. If you are planning a project, your factual questions and answers should help form your plan. If you are analyzing a process for improvement opportunities, your questions and answers should lead to additional questions about possible facts. If you are reviewing a completed project, your factual questions and answers should lead to additional questions about modifying, expanding, or standardizing something.

III. METHODOLOGY

To take full benefit from artifact sharing, research communities and institutions should join efforts to establish common standards for data management and publishing [38]. However, this is a non-trivial task since the expectations toward artifacts vary depending on communities, roles, and artifact types [15].

To design and evaluate a guideline set for quality management of MDE research artifacts, our methodology was informed by five major data sources: (i.) existing general-purpose and discipline-independent quality guidelines of major venues in CS and SE; (ii.) project management literature focusing on quality management [34], [35] and the 5W2H method [35]; (iii.) MDE literature in tooling issues [32] and modeling artifact repositories [29]; (iv.) our own experiences—one author has 10 years of experience in the domain, including membership in artifact evaluation committees (AEC) for 3 MDE-related conferences; (v.) an online survey with MDE experts.

Based on these data sources, we developed a systematic five-phase methodology, schematically depicted in Fig. 1. Our methodology employed the following phases: (i.) identification of practices from guidelines for artifact sharing, (ii.)

categorization of practices based on the 5W2H method for quality management [35], (iii.) definition of factual questions to inquire about practices, (iv.) design and refinement towards a domain-specific guidelines for MDE artifact quality management, and (v.) evaluation and prioritization of the guidelines by MDE experts. Details about each phase are provided in the next subsections. To foster reproducibility and reusability, we made an online supplementary material available on Zenodo [39], Github [40], and in our project website [41].

A. Identification of practices for artifact sharing

To elicitate the quality expectations for research artifacts in the broader field, we analyzed eight guidelines sets for artifact sharing of major CS and SE publishers, venues, and organizations, namely:

- 1) The ACM Artifact Review and Badging [42]
- 2) The EMSE Open Science Initiative [43]–[46]
- 3) The Journal of Open Science Software (JOSS) [47]
- 4) The Journal of Open Research Software (JORS) [48]
- 5) The Guidelines by Wilson et al. [23]
- 6) The NASA Open Source Software Projects [49]
- 7) The TACAS artifact evaluation guideline [50]
- 8) The CAV artifact evaluation guideline [51]

From these guidelines sets, we obtained an initial list of 284 general-purpose practices. This list was compiled in two steps. First, we analyzed each guideline set to extract practices and recommendations for artifact sharing. Second, we refined these extracted practices by standardizing their structure and terminology. In Fig. 2, we show a sample of the practices extracted and their refined versions.

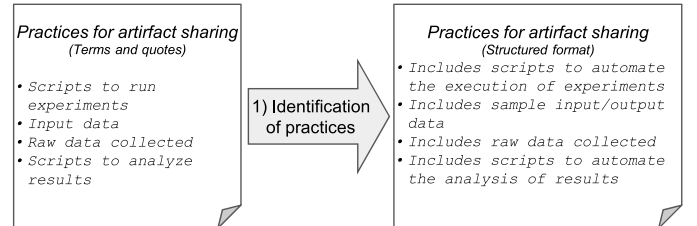


Fig. 2: Examples of extracted and refined practices

B. Categorization of best practices according to the 5W2H

To address the multitude of expectations for artifacts’ usage and quality criteria, we employed the 5W2H method [35] to categorize the extracted practices using the five Ws and two Hs as content tags. We take research practices as *answers to factual questions* that researchers or reviewers could ask about an artifact.

We adapted the 5W2H framework to the context of research artifacts and formulated a pattern to label research practices according to its perspectives. In Table I, we show examples of categorized practices and their respective labels.

TABLE I: Examples of practices and their 5W2H labels

Label	Best practice
What	Provides an indication of the context of the software use
Where	Provide info on how to cite the project (e.g., CITATION file)
When	Provides explanation for changes (e.g., CHANGELOG, commit)
Who	Uses open/non-proprietary file formats
How	There are scripts for every stage of data processing
How	Provides suggestions for other potential applications
How Much	Provides a way to replicate the results with modest resources

The main goal of this step was to gain insights on the types of factual questions that the extracted practices could possibly address. In this task, we employed mind mapping [35] to structure our practices into branches labeled with one of the 5W2H perspectives. We found the three major categories, namely *How*, *What*, and *Where*, composed 83.28% of our initial set of practices. In Table II, we show the percentage of practices categorized in each perspective.

TABLE II: Percentage of practices in each 5W2H perspective

Perspective	%
What	31.44
Where	17.00
Why	3.97
Who	7.93
When	2.83
How	34.84
How Much	1.98

In this labeling process, we used the following pattern: As part of the *What* perspective, we assigned practices associated with the overall description, context and content of the artifact. As part of the *Where* perspective, we assigned practices associated with repository hosting, artifact citation and related work. As part of the *Why* perspective, we assigned practices associated with the reasoning to create an artifact, its objectives and main advantages. As part of the *Who* perspective, we assigned practices associated with usage rights, licensing, authors’ details, and funding agencies. As part of the *When* perspective, we assigned practices associated with version control and identification, updates, and future plans. As part of the *How* perspective, we assigned practices associated with the environment setup, replications, analysis of results, and repurposing. Finally, as part of the *How much* perspective, we assigned practices associated with quantitative information about system requirements and the time needed to run the artifact.

C. Definition of intermediate factual questions

After labeling our practices, we used their associated 5W2H tags to elaborate factual questions that researchers and reviewers could eventually ask about a given research artifact, e.g., “**What** is it about?”, “**Where** shall I cite?”, “**Who** are the authors?”. These questions were designed to help researchers to systematically think about concerns in artifact sharing and provide directions to additional improvement questions. The mind map in Fig. 1 illustrates our factual questions and their associated perspectives. As we discuss in the next section, these questions were designed to kick off the creation of our domain-specific guidelines.

D. Design and refinement of the MDE-specific guidelines

Getting artifacts into publishable shape is often perceived as difficult and time-consuming [52]. To cope with time and resource constraints in research projects, we reviewed our guidelines to identify, match and merge similar practices. After analyzing our initial set of 284 practices, we found various redundant items that were common to different catalogs or covered related issues. Based on these similarities, we elaborated 77 practices to cover one or more items from our initial set of recommendations.

At this point, our guidelines included practices addressing concerns for general types of artifacts, e.g. version control, user instructions, that also apply to MDE artifacts. However, MDE-specific aspects which are known to be important, e.g., model semantics, syntax, and technologies; and influence on the quality of models [30], were still missing.

To tailor our guidelines to the MDE domain, we relied on our own experiences and analyzed two studies from the MDE literature: a taxonomy of tool-related issues affecting the adoption of MDE in the industry [32], and a study on quality criteria for repositories of modeling artifacts [29]. Based on these two studies, we elaborated seven extra practices covering MDE specific concerns and assigned them to an additional factual question inquiring “*What concepts and technologies underpin the artifact?*”. Finally, we incorporated two factual questions which covered associated tasks: “*How to compile/build?*” and “*How to setup a running environment?*”.

This led to our final set of guidelines with 19 factual questions (shown in Fig. 3) and 84 practices. The full set of practices and questions is available in our website [41]. We provide traceability from the 284 analyzed to our 84 practices in our associated artifact (*practices4mde_03Final.pdf* in [39]).

E. Survey

Developing useful research artifacts is challenging as people may have different expectations depending on their role and experience [15]. Thus, we designed a questionnaire survey for MDE experts to study their challenges encountered (RQ2) and to ask them to assess and prioritize our guidelines (RQ3–4). In Table III, we show an overview of our survey.

Participant recruitment. We performed our survey in April-May 2021. Participants were recruited in two main ways: We invited 335 people via e-mail and distributed the invitation on

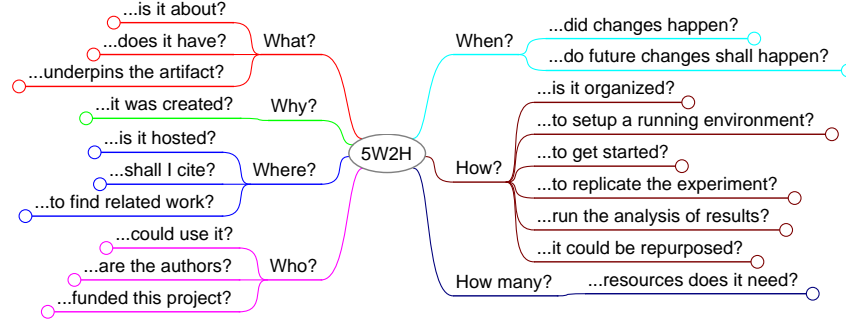


Fig. 3: Mind map with the final set of 19 factual questions

TABLE III: Overview of the survey

Topic	Description
Demographics data	Questions about the participants (Q1) gender and their (Q2) current primary role
General experiences with artifacts	How would you rate your experience in (Q3) artifact development and sharing and (Q4) reusing artifacts in MDE research?; (Q5) Have you ever submitted an artifact for evaluation? Have you ever (Q6) contacted other researchers or (Q7) been contacted by other researchers asking for help on reusing their artifacts?
Challenges in artifact sharing	(Q8) Which challenges have you encountered during the sharing and use of artifacts in MDE research projects?
Evaluation of the Guidelines	We asked participants to rate the (Q9-34) relevance of each one of the 84 practices and, if needed, recommend additional guidelines.
Final evaluation	How do you assess the (Q35) clarity, (Q36) completeness, and (Q37) relevance of these guidelines? Open field for (Q38) additional remarks or (Q39) providing e-mail, if wanted to stay updated about our results.

relevant online channels. The majority of personally invited participants were chosen for having taken a part in the MODELS AEC, coauthored a MODELS paper that earned an ACM artifact badge, and/or coauthored a Software and Systems Modeling (SoSyM) journal paper including some artifact in the last three years. In addition, we invited personal contacts from the MDE community, and encouraged our invitees to forward the invitation to their own contacts with relevant MDE experience. Online channels on which we distributed the call were the PlanetMDE mailing list [53], and our personal LinkedIn and Twitter accounts. Our recruitment activities led to a total of 90 responses.

Questionnaire design. We designed 39 questions to understand the participants’ background and to evaluate the clarity, completeness, and relevance of our guidelines. The questions in our survey covered five topics, as shown in Table III.

First, we collected demographic information, specifically, their gender and current primary role. Demographics are useful for understanding the context of survey’s participants.

Second, we inquired our participants about their general artifact sharing experiences. We asked them to report their level of experience with research artifact sharing and reuse,

both on a 5-points Likert scale. We asked if they have ever submitted an artifact for evaluation and if they have contacted another researcher to reuse artifacts. For the latter, we inquired if they have ever contacted a researcher while trying to reuse an artifact or have ever been contacted by some researcher asking for support with an artifact they previously published. These questions were designed to evaluate the collective experience of our participants with research artifacts.

Third, to gather a domain-specific understanding of what makes MDE artifact sharing and development challenging, we asked the participants to report on challenges encountered during these activities, using an open text field. This field was included to identify issues that could complement our understanding of artifact sharing in MDE research. Based on these responses, we aimed to address **RQ2** by drawing a picture of the challenges faced in MDE artifact sharing and analyzing to what extent our guidelines covered these issues.

Fourth, we asked our participants to prioritize and evaluate our guidelines. We presented all 84 practices to our participants as follows: using the 5W2H perspectives as main categories, each perspective was refined into several factual questions with associated practices proposed as means to address them. We asked our participants to rate each practice as either “Essential”, “Relevant”, or “Unnecessary”, providing a “no answer” option for participants who did not want to rate the practice at hand. To capture any factual questions and practices missed, we also provided an open text field to collect suggestions. These questions were designed to categorize our practices according to their priority and hence, address **RQ3**.

Fifth and finally, we asked our participants to provide an overall score to our guidelines. To this end, we first recapitulated the full set of 19 factual questions. We then asked the participants to evaluate our guidelines considering dimensions: clarity, completeness, and relevance. For each dimension, to obtain a nuanced assessment, participants were asked to specify a score on a 7-point Likert scale. The scale end-points were labeled, in the case of clarity as “very unclear (1)” and “very clear (7)”, and similarly for completeness and relevance.

To collect useful information for interpreting the given scores, we provided an open text field for additional remarks. With this set of scores and additional remarks, we addressed **RQ4**. For participants interested in receiving information about the results of our survey, we provided a text field asking for their e-mail addresses. To counter possible bias, we informed our participants that we would remove the e-mail addresses from our collected data before processing it.

We used the Google Forms platform to perform the survey. In our dry runs, completing the questionnaire took around 15 minutes, which we communicated as an estimate.

IV. RESULTS

We now present our results: our practices and the insights about them brought forward in our MDE expert survey. The presentation of results is organized into five parts. First, we give a brief overview of our practices (addressing RQ1). Second, we discuss our respondents' demographic characteristics and experiences with research artifacts. Third, we analyze the challenges reported by our participants on artifact sharing and reuse, and how our guidelines address them (RQ2). Fourth, we analyze how our participants prioritized our set of practices (RQ3). Finally, we present our participants' assessment of the completeness, clarity, and relevance of our guidelines (RQ4).

A. Overview of our guidelines for MDE artifact sharing (RQ1)

Our guidelines for artifact sharing in MDE research comprise a structured set of 84 best practices. These best practices are proposed as answers to 19 different factual questions (shown in Fig. 3) that researchers may ask about an MDE research artifact. These factual questions cover the seven perspectives of the 5W2H framework and should encourage researchers to inquire about artifact's quality concerns. Table VII shows a selection of our practices based on a prioritization from our survey participants (explained in Sect. IV-D).

While our guidelines aim to cover all relevant aspects of MDE artifact sharing in an encompassing manner, we acknowledge that they might need to be tailored to particular circumstances. For example, for an artifact that is not executable (such as a collection of models), some guidelines in category 6.4) *How to replicate the experiment?* may not be applicable. Users of the guidelines, such as artifact authors and artifact evaluation organizers, should reflect on the guidelines and apply them in a way that is meaningful in their particular circumstances.

Guidelines for MDE artifact sharing (RQ1)

We derived a set of 84 practices, structured along 19 factual questions, to provide guidance to artifact sharing in MDE research. Our guidelines are proposed as means to address factual questions that researchers may inquire about an MDE artifact and visualize "artifact quality" from different perspectives.

B. Survey demographics and experiences with artifacts

Based on our recruitment activities, we obtained a total of 90 responses. In Table IV, we capture an overall picture of our respondents' gender and current primary role.

TABLE IV: Demographics - Primary role and gender

Primary role	# Male	# Female
Industrial Practitioner	7	0
Industrial Researcher	7	0
Academic (Pre-Phd)	10	4
Academic (Post-Doc)	18	5
Academic (Professor)	35	4

In our poll, 43.3% identified themselves as *Academic Professor* and 85.6% were identified as male. While roles were more evenly distributed among *Female academics*, the majority of our male respondents were academic professors. Although we provided an open text field for non-binary genders, no participant used it.

Regarding our participants' experiences with artifact reuse, 83.2% reported having either made contact with another researcher or been contacted by another researcher asking for support with an artifact. These findings indicate that our participants have meaningful collective experiences with research artifacts. In Table V, we show the numbers of participants who have made contact with or been contacted by another researcher.

TABLE V: Have you made contact/contacted someone for the purpose of artifact reuse?

Made contact?	Been contacted?	#	%
No	Yes	13	14.4
Yes	No	13	14.4
No	No	15	16.7
Yes	Yes	49	54.4

C. How do the proposed guidelines address challenges encountered by MDE experts? (RQ2)

In this section, we discuss the top ten challenges reported by MDE experts and how our guidelines cover these issues. To understand how our guidelines address the challenges faced by MDE experts, we first analyzed the answers to an open-ended question we provided to identify "*issues that make the sharing and use of artifacts difficult*". From our 90 participants, we obtained 66 answers that we analyzed, tagged, and compared against our guidelines. The full set of 66 answers is available as supplementary material [39], [40].

1) *What are the challenges encountered?*: To analyze the challenges reported by MDE experts, we employed open coding to classify, group, and quantify answers based on their main concern. One author was responsible for the coding process; the other reviewed the assigned tags. In total, we identified 28 groups of answers from which the top ten challenges are shown in Fig. 4 with their respective identifiers.

The identified challenges largely match those identified for general SE in a previous study [52]; however, there are a number of noteworthy exceptions, such as *technology*

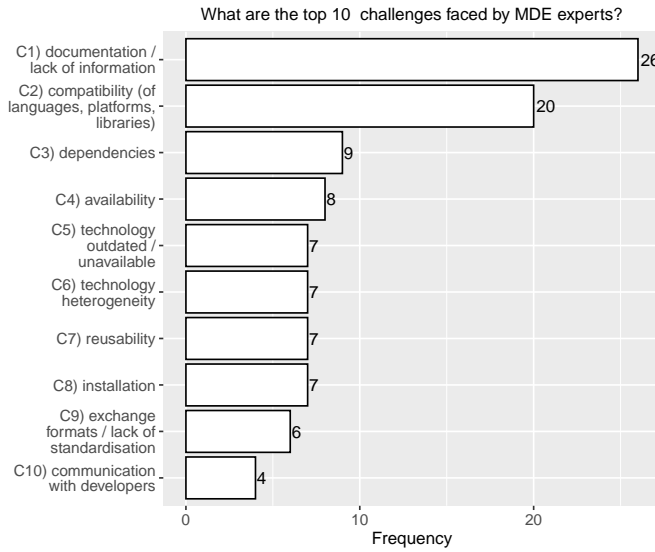


Fig. 4: Top 10 challenges faced in MDE artifact sharing

heterogeneity, data exchange formats, and the lack of standardisation, that can make MDE artifacts more difficult to be reused and produced.

Also corroborating the findings by Timperley et al. [52], we found that *the lack of information and documentation* about artifacts was the topmost challenge faced by our MDE experts. As one participant indicated:

“Textual description about an model can be useful to better explain the model and mitigating doubts.” [P75]

Human comprehension is known to be an important task that contributes towards high-quality modeling artifacts [30]. Thus, to enhance researchers’ comprehension, artifact creators should provide useful *information and documentation* that describe the context of development and relevance of the artifact to the addressed problem, and facilities it offers.

In the second place, we found *compatibility issues* among the topmost reported challenges. To mitigate this problem, artifact authors should always provide details about the technologies and concepts that underpin the artifact. Particularly, these should include the version identifiers of modeling languages, input file formats, or third-party artifacts used in the project, e.g., libraries, frameworks, integrated development environments.

Although reporting a detailed description of an artifact may be seen as irrelevant or not worthy [52], finding a *good-enough* amount of information can improve the quality of artifacts and facilitate their future reuse and repurposing. For example, indicating the operating system and hardware context in which the artifact was developed and tested shall support on setting up experimental environments.

2) *To what extent do the guidelines cover them?:* In this section, we analyze the top ten challenges to identify to what extent they have been covered by our factual questions and their respective practices. In Table VI, we depict a traceability matrix summarizing what perspectives have at least one prac-

TABLE VI: Traceability matrix for the 5W2H perspectives and Top 10 challenges encountered by MDE experts

5W2H	Question	Challenge									
		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
What	1.1) What is it all about?	✓									
	1.2) What does it have?	✓		✓	✓			✓			
	1.3) What underpins the artifact?	✓	✓	✓	✓	✓	✓	✓		✓	
Why	2.1) Why it was created?	✓									
Where	3.1) Where is it hosted?	✓			✓						
	3.2) Where shall I cite?	✓									
	3.3) Where to find related work?	✓									
Who	4.1) Who could use it?	✓						✓		✓	
	4.2) Who are the authors?	✓									✓
	4.3) Who funded this project?	✓									
When	5.1) When did changes happen?	✓			✓			✓			
	5.2) When do future changes shall happen?	✓			✓			✓			✓
How	6.1) How is it organized?	✓		✓				✓		✓	
	6.2) How to setup a running environment?	✓	✓	✓	✓	✓	✓	✓	✓		
	6.3) How to get started?	✓		✓	✓			✓			
	6.4) How to replicate the experiment?	✓			✓			✓			
	6.5) How to run the analysis of results?	✓			✓			✓			
	6.6) How could it be repurposed?	✓	✓	✓	✓	✓	✓	✓			
How many	7.1) How many resources does it need?	✓		✓				✓	✓		

C1=Documentation C2=Compatibility C3=Dependencies C4=Availability C5=Tech. outdated/unavailable
C6=Tech. Heterogeneity C7=Reusability C8=Installation C9=Exchange formats C10=Communication

tice able to cover a given challenge. Challenges are shown by their respective rank identifier and marked cells indicate that perspective-challenge mapping.

Guaranteeing that useful information about an artifact is available is one of the main goals of our guidelines. Hence, in all seven perspectives, we identified at least one practice that could address challenge C1.

In the perspectives *What*, *How* and *How Many*, we identified various practices able to address challenges from C2 to C7. Examples of these practices are *relying on well-maintained libraries*, *reporting known issues/bugs/limitations*, and *indicating of library names and their respective version identifiers*.

Artifact installation is an important step in the practices associated with questions 6.2) and 7.1). Thus, many of their practices were found to be suitable means to cover the challenge C8, such as *providing instructions to install the artifact* and *indicating skills and/or settings required for artifact usage*.

To address challenge C9, we identified practices in questions 1.3), 4.1) and 6.1). Examples of these practices are *reporting standards or specifications used to develop the artifact* and *adopting open/non-proprietary files formats*.

Finally, to address challenge C10, we identified practices in questions 4.2) and 5.2). Examples of these practices are *being open for change requests and receiving feedback from users* and *providing communication channels for interacting with authors and the community*.

Challenges (RQ2)

We identified 28 challenges reported by MDE experts. The two most common challenges by far were a lack of documentation and compatibility issues between languages, platforms, and libraries. By mapping our practices against them, we found that our guidelines provided means to reasonably cover the top 10 challenges.

D. How do MDE experts prioritize the practices? (RQ3)

To evaluate the importance of our proposed guidelines, we analyzed how our participants classified the 84 practices in the three levels of priority: *Essential*, *Desirable*, and *Unnecessary*. These alternatives were adapted from the classification schema used in the ACM SIGSOFT Empirical Standards [10]. In case of doubts or interest in omitting answers, we also provided a *No answer* alternative. Based on the priorities assigned, we identified as *top priority practices* all those which had at least 50% of the participants rating it as an *Essential* item. In Table VII, we show the 23 top priority practices. The full list of all practices with their priorities is available in our supplementary material [39], [40].

We found that six out of the seven 5W2H perspectives had at least one top priority practice. We noticed that the *What* and *How* perspectives showed most of the top priority practices. Particularly, we found that all *What* questions had at least one top priority practice. In the *How* perspective, the only question that did not include any top priority practice was the “6.6) How could it be repurposed?”.

These findings are informative for users of our guidelines, such as artifact developers and organizers of artifact evaluation processes, as we further discuss in Sect. V.

— Prioritization of practices (RQ3) —

Most participants rated most practices at least as desirable, but there is less agreement about the classification into essential vs. desirable. From our 84 practices, we identify a set of 23 top-priority practices that were deemed as essential by more than half of the participants.

E. What is the quality of the proposed guidelines? (RQ4)

In the last part of our questionnaire, we asked participants to provide, on a seven-points Likert scale, an overall score to the completeness, clarity, and relevance of our guidelines. In Fig. 5, we show the frequency of scores for completeness, clarity, and relevance with their respective medians indicated as a vertical dashed line.

Overall, for all three dimensions, more than 92% of our participants reported positive quality scores (5–7 in the seven-point Likert scale). For completeness, clarity, and relevance, we respectively found that 95.5%, 92.2%, and 96.6% of participants reported positive scores. No participant reported a score below three points. In the textual remarks, we found positive sentiments that mirror the positive scores, for example:

“Broad implementation of guidelines such as these is ESSENTIAL for advancing MDE technologies and research!” [P13]

Using the textual feedback, we analyzed the cases in which participants gave negative scores. Considering clarity, which attracted the highest number of negative scores (5.5% of all respondents), we received comments concerning the focus of some practices on artifacts *maintained for a long time* [P33], and the various meanings that “*sharing*” [P58] may take.

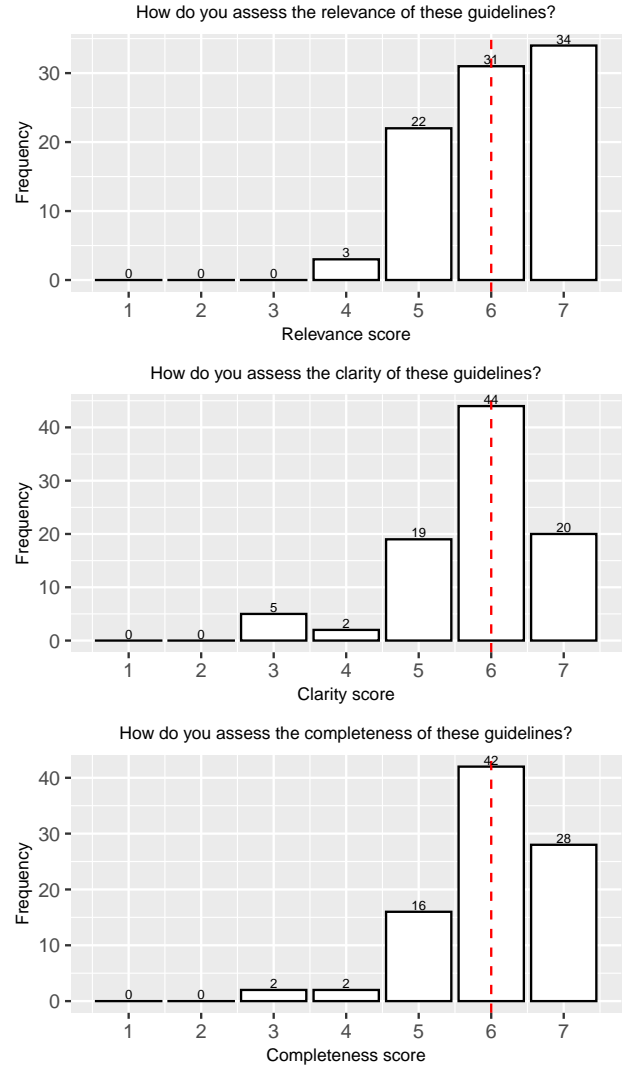


Fig. 5: Relevance, clarity, and completeness ratings

“In terms of clarity, the questions use some abstract terms, in particular sharing. I was somehow confused by this term since sharing MDE artifacts may be associated with a research paper or not. Specially when the artifact is produced in an industrial context.” [P58]

Overall, these findings indicate that our guidelines for MDE artifact sharing were seen as reasonably complete, relevant, and clear. However, we also found there is still room for improvements, such as a need for practices considering special kinds of artifacts, stakeholders, and circumstances in which artifacts may be developed (e.g., industry, academia).

— Evaluated Quality (RQ4) —

The surveyed MDE experts assess the completeness, relevance, and clarity of our guidelines largely positive, with between 92.2% and 96.6% positive scores in each dimension. We identified a number of improvement opportunities.

TABLE VII: Practices for MDE artifact sharing: 23 top-priority practices (out of 84 in total)

5W2H	Question	Practice	Priority		
What	1.1) What is it all about?	Indicate the context of its development (e.g., domain, problem, project)	64.4%	33.3%	2.2%
		Report its name	71.9%	24.7%	3.4%
		Indicate its main functionalities supported (e.g., modeling language, model analysis)	81.1%	18.9%	0.0%
	1.2) What does it have?	Include everything required for replications (i.e., complete)	78.9%	21.1%	0.0%
		1.3) What underpins the artifact?	Indicate modeling languages used to develop it (e.g., UML, SysML, BPMN)	55.1%	40.4%
	Indicate libraries/frameworks used and their respective versions (e.g., Eclipse release)		62.2%	33.3%	4.4%
Why	2.1) Why it was created?	Indicate its objective/goal (e.g., replicability, reusability)	51.1%	44.4%	4.4%
Where	3.1) Where is it hosted?	Repository is open and public (e.g., GitHub, Zenodo, Figshare)	77.8%	17.8%	4.4%
	3.3) Where to find related work?	Give credit to data obtained from other sources (e.g., author, repository)	58.9%	41.1%	0.0%
Who	4.1) Who could use it?	Deposited under an explicit open license (e.g., reported in a LICENSE file)	53.9%	41.6%	4.5%
	4.2) Who are the authors?	Indicate the names of its authors	73.3%	24.4%	2.2%
	4.2) Who are the authors?	Indicate the authors's contact details (e.g., email, ResearchGate, website)	56.7%	36.7%	6.7%
When	5.1) When did changes happen?	Tracked using version control (e.g., GitHub, GitLab, BitBucket)	61.1%	36.7%	2.2%
How	6.1) How is it organized?	Files and folders shall have self-explaining names matching content	51.7%	47.2%	1.1%
		The artifact shall provide a step-by-step tutorial build the source code	60.7%	37.1%	2.2%
	6.2) How to setup a running environment?	The artifact shall provide instructions for downloading	77.5%	18.0%	4.5%
		The artifact shall provide instructions to install it	86.7%	12.2%	1.1%
		The artifact shall include instructions for running it on minimal test data	62.9%	33.7%	3.4%
	6.3) How to get started?	The artifact shall include step-by-step instructions for running it (e.g., README)	86.7%	13.3%	0.0%
		6.4) How to replicate the experiment?	Provide manual/automated instructions for the complete/partial replications	57.3%	41.6%
			The artifact shall include the complete set of test models considered	58.4%	39.3%
How Many	7.1) How many resources does it need?	Provide a clear description of measurements and metrics used in the paper	51.7%	41.6%	6.7%
		Indicate the system/environment settings where it was successfully evaluated	53.3%	42.2%	4.4%
Priority legend: <div>Essential</div> <div>Desirable</div> <div>Unnecessary</div>					

Priority legend: ■ Essential ■ Desirable ■ Unnecessary

V. DISCUSSION

Implications for artifact authors: Our guidelines have been designed as a toolkit to support researchers on the creation, sharing, and maintenance of artifacts in MDE research. The priority levels identified with our survey report what are the most important practices (i.e., essential) so that authors can focus on addressing them. Moreover, the top 10 issues reported by MDE experts can also indicate frequently encountered problems in MDE research and hence, drive the authors' efforts on mitigating them.

Implications for artifact evaluation organizers and reviewers: As previous surveys have shown [15], [54], the lack of consensus on quality standards and discipline-specific guidelines for RDM and artifact sharing opens the opportunity to subjective notions of artifact quality. Our work fills this gap by complementing initiatives, such as the ACM SIGSOFT Empirical Standards, by providing MDE-specific guidelines that can also be used by artifact reviewers in MDE conferences. Since we did not find a clear agreement between participants on the prioritization, our guidelines in their current form are not intended to represent "The" definite list of best practices. However, they can certainly be useful to kick off the creation of venue-specific recommendations, quality criteria, or *frequently asked questions* for MDE research artifacts.

Improvement opportunities: Based on the feedback of our participants, we found there is still room for improvement in our guidelines and noted a few possible improvements. First, employing privacy-preserving techniques, such as software obfuscation [55], before sharing artifacts could be incorporated as an extraordinary industrial research practice. This could foster the disclosure of real-world artifacts in MDE industrial research. Second, to address the lack of viewpoint-specific practices, interviews with different stakeholders (e.g., artifact users, open-source contributors, AEC reviewers, industrial

practitioners/researchers) could be done to understand which particular needs and expectations these actors may have. The interviews could provide insights for building personas and narratives about artifact development, sharing and reuse.

A. Threats to validity

We follow the recommendations by Wohlin et al. [24] to discuss threats to validity. Conclusion validity is out of scope, as we did not search for statistical relationships.

External validity: These concern the generalization of our results to the overall MDE research community. The majority of our participants was formed by male academic professors. Although this may pose threats to the external validity, our findings are still relevant as professors not only have access to their own experiences but to those of their supervised team members as well. One participant explicitly reported on the experiences encountered by his students. These results can also be considered as complementary to other surveys [15] which focused on relatively new scientific peer reviewers, as most of our sample was formed by professors. Another variable that may form a threat to external validity is the small female and non-binary participation. Women's participation in open science has been increasing over time [56] and hence may have specific influences in artifact sharing.

Internal validity: These threats concern issues that may indicate a causal relationship when there is none. As the validity of surveys is highly dependent on its audience's representativity, we tried to cover the MDE community via two main channels: the PlanetMDE mailing list [53] and by approaching authors of recent SoSyM and MODELS papers. Another variable that may constitute a threat to internal validity is the social-desirability bias [57]. To mitigate this bias, we made our questionnaire anonymous to avoid the identification of participants, following Wohlin et al.'s recommendation [24]. Participants

could still opt-in to receive our results by providing an email address, but were informed that we would remove the e-mail addresses from our data before analyzing them.

Construct validity: These are concerned with the ability to draw correct conclusions about the treatment and outcomes. To mitigate this kind of threat, we focused our survey on MDE experts which should be expected to understand the specificities of MDE artifacts. Moreover, to identify potential misunderstandings, we provided means to the participants to explain their choices. From our analysis, it seems that there were no major sources of misunderstanding, which is also in line with the largely positive clarity scores obtained for RQ4.

VI. RELATED WORK

Artifact sharing: Hermann et al. [15] surveyed AEC members of major CS conferences and found that the community expectations of artifact quality exceed the ones expressed in calls for artifacts and reviewing guidelines. Additionally, they found there is no consensus on quality thresholds for research artifacts in general. Heumüller et al. [54] analyzed ICSE papers published from the years 2007 to 2017 and identified a positive trend towards artifact availability and a small, but statistically significant, positive correlation between linking artifacts to a paper and its citations. Timperley et al. [52] reported several high-level challenges that affect the quality of artifacts and mismatched expectations between artifact creators, users and reviewers. Our paper complements this literature by introducing a domain-specific quality guideline set and investigating the opinion of domain experts about our proposed guidelines.

The 5W2H framework: Jia et al. [58] proposed and reported their experiences with a 5W+1H pattern to examine systematic mapping studies from a generic set of dimensions. Their pattern is proposed as a tool for investigators to define a set of systematic, generic, and complementary RQs, enabling them to kick off and expedite the mapping study process in a well-defined manner. Prana et al. [59] investigated the problem of automated classification of README file content. Using the 5W2H framework to manually annotate README file sections, the authors show that their approach can support repository owners to improve the quality of software documentation. Zhang et al. [60] proposed an approach to generate automatic summarization of scientific literature based on 5W1H event structure and trigger word templates. Compared with existing abstracts given by authors, their approach was able to provide more detailed information, in a more convenient format. In our methodology, we manually categorized research practices as answers to factual questions following the 5W2H framework. However, the process for identifying research practices could be still automated, at least partially, by means of identifying trigger word templates for research practices.

Research Data Management: Perrier et al. [61] conducted a scoping review of RDM in academic institutions and found that studies investigating processes to improve the quality of data could potentially provide tangible guidance to researchers

interested in effective data reuse. Van Eeuwijk et al. [26] noted that research software is a multi-faceted asset and very diverse in terms of size, complexity and format. Hence, software sustainability policies should reflect the characteristics of different software and research domains. Marjan et al. [14] surveyed researchers from Horizon 2020 projects and identified a need for much more tailored guidance and domain-specific standards examples of data management plans. In this work, we fill these gaps by tailoring general practices for artifact sharing to the MDE research domain. The principles underpinning our investigation can be extended to other domains.

VII. CONCLUSION AND FUTURE WORK

Artifact sharing is known to be helpful for researchers and practitioners to build upon existing knowledge, adopt novel contributions in practice, and increase the chances of papers receiving citations. In MDE research, there is an urge for artifact sharing as the community targets a broader use of AI-based techniques, which can only become feasible if large open datasets and confidence measures about their quality are available. In this paper, we introduce a set of quality guidelines specifically tailored for MDE research artifacts.

Based on project management principles, we designed a catalog of 84 MDE-specific research recommendations from generic practices for artifact sharing and domain-specific literature about MDE tooling issues and modeling artifact repositories. These practices are proposed as answers to factual questions that researchers can use to systematically think about concerns in MDE artifact sharing and provide directions to additional improvement inquiring.

In a poll among 90 MDE experts, more than 92% positively assessed the clarity, completeness, and relevance of our guidelines. Our participants reported priority levels to our practices which can guide the research decision-making during the creation, sharing, reuse, and evaluation of MDE research artifacts. The full set of generic practices, MDE-specific guidelines, and factual questions are provided as supplementary material [39]–[41]. Particularly, we highlight our project website [41] available at <https://mdeartifacts.github.io/> which can be used by artifact authors, researchers, and AECs of MDE conferences and journals.

There are several relevant directions for future work. First, our guidelines could still be improved by indicating in which context (i.e., artifact type) and for whom (viewpoints) a given practice should be seen as essential, desirable, or unnecessary. Second, to determine the effect of aspects such as gender and previous experiences on the prioritization of guidelines, it would be interesting to perform sub-group analysis on our data. Finally, our methodology focused exclusively on MDE artifacts, but it could be easily extended to other domains. Experimenting with our methodology in other domains such as software product line engineering, in which a need for consolidated community benchmarks has been expressed [62], would be a desirable contribution.

REFERENCES

- [1] H. Pashler and E.-J. Wagenmakers, “Editors’ introduction to the special section on replicability in psychological science: A crisis of confidence?” *Perspectives on Psychological Science*, vol. 7, no. 6, pp. 528–530, 2012, place: US Publisher: Sage Publications.
- [2] C. Collberg and T. A. Proebsting, “Repeatability in computer systems research,” *Communications of the ACM*, vol. 59, no. 3, pp. 62–69, Feb. 2016.
- [3] J. Lung, J. Aranda, S. M. Easterbrook, and G. V. Wilson, “On the difficulty of replicating human subjects studies in software engineering,” in *Proceedings of the 30th international conference on Software engineering*, ser. ICSE ’08. New York, NY, USA: Association for Computing Machinery, May 2008, pp. 191–200.
- [4] L. Glanz, S. Amann, M. Eichberg, M. Reif, B. Hermann, J. Lerch, and M. Mezini, “CodeMatch: obfuscation won’t conceal your repackaged app,” in *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering*, ser. ESEC/FSE 2017. New York, NY, USA: Association for Computing Machinery, Aug. 2017, pp. 638–648.
- [5] ESEC/FSE, “Call for Artifact Evaluation | ESEC/FSE 2011,” 2011. [Online]. Available: <http://2011.esec-fse.org/cfp-artifact-evaluation>
- [6] —, “ESEC/FSE 2020 - Artifacts - ESEC/FSE 2020,” 2020. [Online]. Available: <https://2020.esec-fse.org/track/esecfse-2020-artifacts>
- [7] ICSE, “ICSE 2020 - Artifact Evaluation - ICSE 2020,” 2020. [Online]. Available: <https://2020.icse-conferences.org/track/icse-2020-Artifact-Evaluation>
- [8] SPLC, “Call for research papers – 24th ACM International Systems and Software Product Line Conference,” 2020. [Online]. Available: <https://splc2020.net/call-for-papers/call-for-research-papers/>
- [9] MODELS, “MODELS 2020 - Artifact Evaluation - MODELS 2020,” 2020. [Online]. Available: <https://conf.researchr.org/track/models-2020/models-2020-artifact-evaluation>
- [10] P. Ralph, S. Baltés, D. Bianculli, Y. Dittrich, M. Felderer, R. Feldt, A. Filieri, C. A. Furia, D. Gaziotin, P. He, R. Hoda, N. Juristo, B. Kitchenham, R. Robbes, D. Mendez, J. Moller, D. Spinellis, M. Staron, K. Stol, D. Tamburri, M. Torchiano, C. Treude, B. Turhan, and S. Vegas, “ACM SIGSOFT Empirical Standards,” *arXiv:2010.03525 [cs]*, Oct. 2020.
- [11] V. Basili, F. Shull, and F. Lanubile, “Building knowledge through families of experiments,” *IEEE Transactions on Software Engineering*, vol. 25, no. 4, pp. 456–473, Jul. 1999.
- [12] I. von Nostitz-Wallwitz, J. Krüger, and T. Leich, “Towards improving industrial adoption: the choice of programming languages and development environments,” in *Proceedings of the 5th International Workshop on Software Engineering Research and Industrial Practice*, ser. SER&IP’18. New York, NY, USA: Association for Computing Machinery, May 2018, pp. 10–17.
- [13] G. Colavizza, I. Hrynaskiewicz, I. Staden, K. Whitaker, and B. McGillivray, “The citation advantage of linking publications to research data,” *PLOS ONE*, vol. 15, no. 4, p. e0230416, Apr. 2020, arXiv: 1907.02565.
- [14] M. Grootveld, E. Leenarts, S. Jones, E. Hermans, and E. Fankhauser, “OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans,” Jan. 2018. [Online]. Available: <http://doi.org/10.5281/zenodo.1120245>
- [15] B. Hermann, S. Winter, and J. Siegmund, “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*. New York, NY, USA: Association for Computing Machinery, Nov. 2020.
- [16] R. France, J. Bieman, and B. H. Cheng, “Repository for model driven development (remodd),” in *International Conference on Model Driven Engineering Languages and Systems*. Springer, 2006, pp. 311–317.
- [17] O. Babur, “A labeled ecore metamodel dataset for domain clustering,” 2019, type: dataset. [Online]. Available: <https://zenodo.org/record/2585456>
- [18] G. Robles, T. Ho-Quang, R. Hebig, M. R. Chaudron, and M. A. Fernandez, “An extensive dataset of UML models in GitHub,” in *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*. IEEE, 2017, pp. 519–522.
- [19] B. Karasneh, “An online corpus of UML design models: construction and empirical studies,” Ph.D. dissertation, Leiden University, 2016.
- [20] eclipse.org, “ATL Transformation Zoo.” [Online]. Available: <https://www.eclipse.org/atl/atlTransformations/>
- [21] D. Strüber, T. Kehr, T. Arendt, C. Pietsch, and D. Reuling, “Scalability of model transformations: Position paper and benchmark set,” in *Big-MDE’16: Workshop on Scalability in Model-Driven Engineering*, 2016, pp. 21–30.
- [22] F. Basciani, J. Di Rocco, D. Di Ruscio, L. Iovino, and A. Pierantonio, “Model repositories: Will they become reality?” in *CloudMDE@MoDELS*, 2015, pp. 37–42.
- [23] G. Wilson, J. Bryan, K. Cranston, J. Kitzes, L. Nederbragt, and T. K. Teal, “Good enough practices in scientific computing,” *PLOS Computational Biology*, vol. 13, no. 6, p. e1005510, Jun. 2017, publisher: Public Library of Science.
- [24] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslin, *Experimentation in Software Engineering*. Springer Publishing Company, Incorporated, 2012.
- [25] R. University, “Research Data management,” 2021, last Modified: 2019-06-14.
- [26] S. van Eeuwijk, T. Bakker, M. Cruz, V. Sarkol, B. Vreede, B. Aben, P. Aerts, G. Coen, B. van Dijk, P. Hinrich, L. Karvovskaya, M. Keijzer-de Ruijter, J. Koster, J. Maassen, M. Roelofs, J. Rijnders, A. Schrotten, L. Sesink, C. van der Togt, J. Vinju, and P. de Willigen, “Research software sustainability in the Netherlands: Current practices and recommendations,” Zenodo, Tech. Rep., Feb. 2021.
- [27] L. Corti, V. Van den Eynden, L. Bishop, and M. Woollard, *Managing and sharing research data: a guide to good practice*, 2nd ed. Thousand Oaks, CA: SAGE Publications, 2019.
- [28] M. Brambilla, J. Cabot, and M. Wimmer, *Model-driven software engineering in practice*, ser. Synthesis lectures on software engineering. San Rafael, Calif.: Morgan & Claypool, 2012, no. 1, oCLC: 820461802.
- [29] F. P. Basso, C. M. L. Werner, and T. C. Oliveira, “Revisiting Criteria for Description of MDE Artifacts,” in *2017 IEEE/ACM Joint 5th International Workshop on Software Engineering for Systems-of-Systems and 11th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (JSOS)*, May 2017, pp. 27–33.
- [30] J. Krogstie, “Quality of Models,” in *Model-Based Development and Evolution of Information Systems: A Quality Approach*, J. Krogstie, Ed. London: Springer, 2012, pp. 205–247.
- [31] D. Steinberg, F. Budinsky, M. Paternostro, and E. Merks, *EMF: Eclipse Modeling Framework, 2nd Edition*, 2nd ed. Addison-Wesley Professional, Dec. 2008.
- [32] J. Whittle, J. Hutchinson, M. Rouncefield, H. Burden, and R. Heldal, “A taxonomy of tool-related issues affecting the adoption of model-driven engineering,” *Software & Systems Modeling*, vol. 16, no. 2, pp. 313–331, May 2017.
- [33] R. Katz, “Challenges in Doctoral Research Project Management: A Comparative Study,” *International Journal of Doctoral Studies*, vol. 11, pp. 105–125, 2016.
- [34] P. M. Institute, Ed., *A guide to the project management body of knowledge / Project Management Institute*, 6th ed., ser. PMBOK guide. Newtown Square, PA: Project Management Institute, 2017.
- [35] N. R. Tague, *The quality toolbox*, 2nd ed. Milwaukee, Wis: ASQ Quality Press, 2005.
- [36] Z. Pan and G. M. Kosicki, “Framing analysis: An approach to news discourse,” *Political Communication*, vol. 10, no. 1, pp. 55–75, Jan. 1993.
- [37] G. Hart, “The five Ws: An old tool for the new task of audience analysis,” *Technical Communication*, vol. 43, no. 2, pp. 139–145, 1996. [Online]. Available: <http://www.geoff-hart.com/articles/1995-1998/five-w.htm>
- [38] The Royal Society, “Science As an Open Enterprise,” Royal Society, London, Tech. Rep., 2012. [Online]. Available: <https://royalsociety.org/-/media/policy/projects/sape/2012-06-20-saoe.pdf>
- [39] C. D. N. Damasceno and D. Strüber, “damascenodiego/mdeartifacts.github.io: Artifacts for this paper,” Jul. 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.5109401>
- [40] —, “damascenodiego/mdeartifacts.github.io,” Jul. 2021. [Online]. Available: <https://github.com/damascenodiego/mdeartifacts.github.io>
- [41] —, “The MDE Artifacts project,” 2021. [Online]. Available: <https://mdeartifacts.github.io/>
- [42] ACM, “Artifact Review and Badging - Current,” Aug. 2020. [Online]. Available: <https://www.acm.org/publications/policies/artifact-review-and-badging-current>
- [43] D. Méndez Fernández, M. Monperrus, R. Feldt, and T. Zimmermann, “The open science initiative of the Empirical Software Engineering

- journal,” *Empirical Software Engineering*, vol. 24, no. 3, pp. 1057–1060, Jun. 2019.
- [44] M. Monperrus, “How to make a good open-science repository?” Dec. 2019, section: Updates in Data. [Online]. Available: <https://researchdata.springernature.com/posts/57389-how-to-make-a-good-open-science-repository>
- [45] EMSE, “EMSE Open Science Initiative,” Mar. 2021, original-date: 2018-06-28T15:35:56Z. [Online]. Available: <https://github.com/emsejournal/openscience/blob/master/README.md>
- [46] EMSE, “EMSE Open science - Evaluation Criteria,” Mar. 2021. [Online]. Available: <https://github.com/emsejournal/openscience/blob/master/review-criteria.md>
- [47] D. S. Katz, K. E. Niemeyer, and A. M. Smith, “Publish your software: Introducing the Journal of Open Source Software (JOSS),” *Computing in Science Engineering*, vol. 20, no. 3, pp. 84–88, May 2018, conference Name: Computing in Science Engineering.
- [48] JORS, “The Journal of Open Research Software - Editorial Policies,” Feb. 2021. [Online]. Available: <http://openresearchsoftware.metajnl.com/about/editorialpolicies/>
- [49] NASA, “NASA Open Source Software,” 2021. [Online]. Available: <https://code.nasa.gov/>
- [50] TACAS, “TACAS 2019 - ETAPS 2019,” 2019. [Online]. Available: <https://conf.researchr.org/track/etaps-2019/tacas-2019-papers#Artifact-Evaluation>
- [51] CAV, “Artifacts | CAV 2019,” 2019. [Online]. Available: <http://i-cav.org/2019/artifacts/>
- [52] C. S. Timperley, L. Herckis, C. Le Goues, and M. Hilton, “Understanding and improving artifact sharing in software engineering research,” *Empirical Software Engineering*, vol. 26, no. 4, p. 67, 2021.
- [53] PlanetMDE. planetmde - model driven engineering ANNOUNCEMENTS - info. [Online]. Available: <https://listes.univ-grenoble-alpes.fr/sympa/info/planetmde>
- [54] R. Heumüller, S. Nielebock, J. Krüger, and F. Ortmeier, “Publish or perish, but do not forget your software artifacts,” *Empirical Software Engineering*, vol. 25, no. 6, pp. 4585–4616, Nov. 2020.
- [55] M. Jasper, M. Mues, A. Murtovi, M. Schlüter, F. Howar, B. Steffen, M. Schordan, D. Hendriks, R. Schiffelers, H. Kuppens, and F. W. Vaandrager, “Rers 2019: Combining synthesis with real-world models,” in *Tools and Algorithms for the Construction and Analysis of Systems*, D. Beyer, M. Huisman, F. Kordon, and B. Steffen, Eds. Cham: Springer International Publishing, 2019, pp. 101–115.
- [56] M. C. Murphy, A. F. Mejia, J. Mejia, X. Yan, S. Cheryan, N. Dasgupta, M. Destin, S. A. Fryberg, J. A. Garcia, E. L. Haines, J. M. Harackiewicz, A. Ledgerwood, C. A. Moss-Racusin, L. E. Park, S. P. Perry, K. A. Ratliff, A. Rattan, D. T. Sanchez, K. Savani, D. Sekaquaptewa, J. L. Smith, V. J. Taylor, D. B. Thoman, D. A. Wout, P. L. Mabry, S. Ressler, A. B. Diekman, and F. Pestilli, “Open science, communal culture, and women’s participation in the movement to improve science,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 39, pp. 24 154–24 164, Sep. 2020.
- [57] Wikipedia, “Social-desirability bias,” 2020, page Version ID: 992112847. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Social-desirability_bias&oldid=992112847
- [58] C. Jia, Y. Cai, Y. T. Yu, and T. H. Tse, “5W+1H pattern: A perspective of systematic mapping studies and a case study on cloud software testing,” *Journal of Systems and Software*, vol. 116, pp. 206–219, Jun. 2016.
- [59] G. A. A. Prana, C. Treude, F. Thung, T. Atapattu, and D. Lo, “Categorizing the Content of GitHub README Files,” *Empirical Software Engineering*, vol. 24, no. 3, pp. 1296–1327, Jun. 2019.
- [60] J. Zhang, K. Li, C. Yao, and Y. Sun, “Event-based summarization method for scientific literature,” *Personal and Ubiquitous Computing*, Apr. 2020.
- [61] L. Perrier, E. Blondal, A. P. Ayala, D. Dearborn, T. Kenny, D. Lightfoot, R. Reka, M. Thuna, L. Trimble, and H. MacDonald, “Research data management in academic institutions: A scoping review,” *PLOS ONE*, vol. 12, no. 5, p. e0178261, May 2017, publisher: Public Library of Science.
- [62] D. Strüber, M. Mukelabai, J. Krüger, S. Fischer, L. Linsbauer, J. Martinez, and T. Berger, “Facing the truth: benchmarking the techniques for the evolution of variant-rich systems,” in *SPLC’19: International Systems and Software Product Line Conference*. ACM, 2019, pp. 26:1–10.

Quality Guidelines for Research Artifacts in Model-Driven Engineering (Artifact Paper)

Carlos Diego Nascimento Damasceno, and Daniel Strüber

Institute for Computing and Information Sciences

Radboud University, Nijmegen, Netherlands

{d.damasceno,d.strueber}@cs.ru.nl

Abstract—Sharing research artifacts is known to help people to build upon existing knowledge, adopt novel contributions in practice, and increase the chances of papers receiving attention. In Model-Driven Engineering (MDE), openly providing research artifacts plays a key role, even more so as the community targets a broader use of AI techniques, which can only become feasible if large open datasets and confidence measures for their quality are available. However, the current lack of common discipline-specific guidelines for research data sharing opens the opportunity for misunderstandings about the true potential of research artifacts and subjective expectations regarding artifact quality. In this artifact paper, we describe the research artifacts created from the homonym paper [1] accepted in the Foundation Track of the MoDELS’21. To foster the reproducibility and reusability of our results, we make the artifact of this study available on the website: <https://mdeartifacts.github.io/>.

Index Terms—Software artifacts, Open Science, Model-Driven Engineering, Quality Management

I. WHAT THIS ARTIFACT DOES/INCLUDES?

In this artifact paper, we describe the artifacts created from the paper **Quality Guidelines for Research Artifacts in Model-Driven Engineering** [1] accepted in the MoDELS’21. In this artifact, you will find: (1) Text documents (in *.docx* format) with the full set of research practices extracted; (2) Various mind maps (in *.mm* format) that we created to categorize research practices according to the 5W2H framework; (3) Our guidelines, which include the 19 factual questions and 84 best practices specifically tailored to MDE research projects; (4) Our Survey questionnaire and responses (between April-May 2021) and R scripts for analyzing the responses and drawing the plots shown in the paper.

II. WHERE IT CAN BE FOUND?

This artifact is hosted on GitHub as an open-source project, under the MIT license [2]. For long-term availability, we also archived our project on Zenodo [3] Finally, we also plan to maintain a website for this project [4].

III. HOW TO REPRODUCE THIS WORK?

A successful replication/reproduction of our work entails the following: (1) Curation of research practices common to a given research community; (2) Classification of practices according to the 5w2h method; (3) Refinement of practices (e.g., remove redundancy, addition of missing practices); (4) Repeating the survey among experts in the research community and the analysis of responses. To design your own replication/reproduction, we recommend the interested reader to look at the files available in the directory `artifacts/202107_models/` of this artifact [2].

In the directory `202107_models/data/`, you will find the *.docx* files and mind maps in *.mm* format with the practices and see how those could be organized. To open the mind maps in *.mm* format, we recommend the user to use the FreePlane Mind Mapping tool [5], which is multiplatform.

In the directory `202107_models/analysis/`, we also provide a set of RScripts [6] and an RStudio project [7] for reproducing the plots shown in the paper, e.g., participants’ role, challenges faced by MDE experts, overall satisfaction of participants, top-priority practices. To reproduce these steps, please check the R scripts `analysis/analysis_0[1-4]*.Rmd`. These files shall reproduce the analysis of the survey responses and figures we used to answer the Research Questions 1 to 4.

IV. ACM REVIEWING AND BADGING

We believe this artifact fulfills the criteria specified in the document *ACM Review and Badging v1.1* [8], particularly for the badges **Artifacts Available** and **Artifacts Evaluated – Reusable**. For the Artifacts Available, we have given the location of the artifact on GitHub [2] and a DOI issued by Zenodo [3]. For the Artifacts Evaluated – Reusable, we have made an effort to provide documentation explaining how to obtain the artifact, what are the main parts of the artifact, how to get started, and how to use the artifact in more detail.

V. FINAL REMARKS

As an open-source initiative, we are happy to have your collaboration and contributions towards better MDE artifacts. Feel free to contact us via GitHub or email!

REFERENCES

- [1] C. D. N. Damasceno and D. Strüber, “Quality guidelines for research artifacts in model-driven engineering,” in *MoDELS’21: ACM/IEEE 24th International Conference on Model Driven Engineering Languages and Systems, Virtual Event, Japan, 10-15 October, 2021*. ACM, 2021.
- [2] —, “damascenodiego/mdeartifacts.github.io,” Jul. 2021. [Online]. Available: <https://github.com/damascenodiego/mdeartifacts.github.io>
- [3] —, “damascenodiego/mdeartifacts.github.io: Artifacts used in the paper accepted at the MODELS Conference,” Jul. 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.5109401>
- [4] —, “The MDE Artifacts project,” 2021. [Online]. Available: <https://mdeartifacts.github.io/>
- [5] Freeplane, “Freeplane - free mind mapping and knowledge management software.” [Online]. Available: <https://www.freeplane.org/>
- [6] The R Project, “R: The R Project for Statistical Computing.” [Online]. Available: <https://www.r-project.org/>
- [7] RStudio, “Download the RStudio IDE.” [Online]. Available: <https://rstudio.com/products/rstudio/download/>
- [8] ACM, “Artifact Review and Badging - v.1.1,” Aug. 2020, [Online].