
Boas práticas para gerenciamento de qualidade de artefatos de pesquisa em engenharia de software

Carlos Diego Nascimento Damasceno^{1*}; Isotilia Costa Melo²

¹ Radboud University. Department of Software Science. Pesquisador doutor. Toernooiveld 212; 6525 EC Nijmegen, The Netherlands

² Universidad Adolfo Ibáñez (UAI). Pesquisadora Doutora. Padre Hurtado 750 – Miraflores; 25.803-35 Viña del Mar, Valparaíso, Chile

*autor correspondente: d.damasceno@cs.ru.nl

Boas práticas para gerenciamento de qualidade de artefatos de pesquisa em engenharia de software

Resumo

O compartilhamento de artefatos de pesquisa em engenharia de software (ES) é uma conhecida atividade capaz de fortalecer a transparência de estudos científicos, alavancar o desenvolvimento de novas tecnologias e incentivar profissionais da indústria a incorporar inovações tecnológicas em seus portfólios de ferramentas. Entretanto, a carência de ferramentas especificamente projetadas para guiar o desenvolvimento e compartilhamento de artefatos de pesquisa em ES constitui uma ameaça para a reprodutibilidade e sustentabilidade de projetos de pesquisas. Para mitigar estes problemas, este trabalho tem como objetivo principal apresentar um processo de desenvolvimento e validação de um conjunto de boas práticas para auxiliar no gerenciamento de qualidade de artefatos de pesquisa em engenharia de software dirigida por modelos, uma variante de ES que usa modelos abstratos para guiar o projeto e desenvolvimento de sistemas de software. Em um questionário realizado com 90 especialistas no tema, o conjunto de práticas proposto foi avaliado positivamente por de cerca de 92% dos seus participantes. Para ampliar o alcance deste conjunto de práticas como ferramenta de gerenciamento de qualidade de artefatos de pesquisa, as contribuições discutidas neste trabalho encontram-se disponibilizadas em dois repositórios de versionamento e arquivamento largamente usados em ciência aberta.

Palavras-chave: Gerenciamento de qualidade; Engenharia de Software; Pesquisa Científica; Artefatos de Software.

Best practices for artifact quality management in software engineering research

Abstract

Artifact sharing in software engineering (SE) research is a practice known for strengthening the transparency of scientific studies, aiding researchers in the development of new technologies, and encouraging practitioners to incorporate technological innovations into their portfolios. However, the current lack of methodologies specifically designed to guide the development and sharing of research artifacts poses a threat to the reproducibility and sustainability of SE research projects. To mitigate these risks, this work has as main objective to introduce a process for designing and evaluating a set of best practices to assist in the quality management of research artifacts in model-driven software engineering, a SE variant that relies on abstract models to guide the project and development of software systems. Through a questionnaire carried out with 90 experts on the subject, the proposed set of best practices was positively evaluated by about 92% of its participants. To expand the reach of this set of practices, the contributions discussed in this paper are also made available in two versioning and archiving repositories widely used in open science.

Keywords: Quality management; Software engineering; Scientific research; Software artifacts.

Introdução

Pesquisas em engenharia de software [ES] têm como objetivo estudar e desenvolver conceitos e técnicas que auxiliem na criação, operação e manutenção de sistemas e projetos de software. Complementando o papel de artigos de revistas e conferências científicas, artefatos de pesquisa vêm sendo reconhecidos como um importante produto para a disseminação de resultados de projetos de pesquisa científica e para aumentar a reprodutibilidade e reusabilidade de estudos científicos (Heumüller et al., 2020). Adicionalmente, artefatos de pesquisa atuam de forma complementar, ou até mesmo substituir, as outras formas de experimentação, tais como “in vivo” e “in vitro”.

Artefatos de pesquisa científica podem ter diferentes tipos e propósitos, tais como “scripts” e tabelas de dados para análises estatísticas, protótipos para prova-de-conceito, e ferramentas de software completas para uso diário ou na resolução de problemas específicos em ambientes de produção (Heumüller et al., 2020). Neste trabalho, considera-se como artefato de pesquisa qualquer tipo de objeto digital que tenha sido diretamente criado pelos autores de um estudo científico ou gerado como resultado de um experimento científico (ACM, 2020). Artefatos de pesquisa podem auxiliar pesquisadores e profissionais da indústria no desenvolvimento de novas metodologias (Hermann et al., 2020), alavancar a adoção de novas tecnologias, proporcionar uma maior visibilidade e aumentar número de citações de artigos científicos, caso estes incluam referências aos seus respectivos artefatos e dados de pesquisa.

O compartilhamento de artefatos (tais como código fonte, dados de pesquisa e executáveis) é tido como uma prática desejável pela comunidade científica. Atualmente, esta prática é considerada padrão-ouro em pesquisas auxiliadas por computador por favorecer a validação de resultados de estudos científicos, potencializar a visibilidade de estudos e conseqüentemente, sua citação; e fomentar o reuso e repropósito de derivados de pesquisas científicas.

Em reconhecimento a estes benefícios, a comunidade de ES vem incentivando autores de trabalhos aceitos em importantes conferências da área a co-submeterem artefatos associados aos seus artigos a processos complementares de revisão por pares. Iniciativas como essa, alinham-se ao objetivo da comunidade científica e de agências de fomento de desenvolver projetos de pesquisa cada vez mais sustentáveis do ponto de vista técnico, isto é, onde artefatos sejam produzidos de modo que eles possam ser facilmente reutilizados e adaptados em pesquisas futuras (Penzenstadler e Femmer, 2013). Para isso, é necessário que a natureza distinta dos artefatos de pesquisas seja reconhecida como tal e que iniciativas visando a produção de artefatos de qualidade sejam desenvolvidas em comum acordo com a

comunidade científica. Ainda que a autoria e o livre compartilhamento de artefatos de pesquisa sejam conhecidos pela comunidade científica de ES como *bem comum*, estudos mostram que ainda há pesquisadores que não reconhecem o custo-benefício de se criar e compartilhar artefatos de alta-qualidade como algo válido. Além disso, pesquisadores e profissionais da indústria podem frequentemente apresentar expectativas subjetivas sobre o que é *qualidade de artefato de pesquisa* (Van den Eynden et al., 2011), o que pode gerar conflitos durante revisões por pares de artefatos de pesquisa. Para mitigar estes problemas, o estabelecimento e adoção de listas de boas práticas durante o processo de avaliação e revisão de artefatos é visto como benéfico pela comunidade científica (Marjan Grootveld et al., 2018).

Especificamente em Engenharia de Software Dirigida por Modelos, i.e., “Model-Driven Engineering” [MDE], existe uma carência de artefatos de pesquisa e repositórios de modelos de alta-qualidade e complexidade, o que dificulta a realização de estudos empíricos robustos e “benchmarking” de técnicas de MDE (Basciani et al., 2019). Dessa forma, o desenvolvimento de um catálogo de recomendações e boas práticas de pesquisa em MDE é bem-vindo pela comunidade de pesquisadores e profissionais da indústria, uma vez que há um crescente interesse em soluções de MDE de uso intensivo de dados, tais como baseadas em ciências de dados e aprendizado de máquina.

Este trabalho tem como objetivo apresentar o processo de desenvolvimento e validação de um conjunto de boas práticas de pesquisa para auxiliar no gerenciamento de qualidade de artefatos de pesquisa em MDE.

Material e Métodos

Para desenvolver um catálogo de boas práticas de compartilhamento de artefatos de pesquisa em MDE, foram usados princípios de gerenciamento de qualidade de projeto, em particular do processo de Planejar o Gerenciamento de Qualidade (PMI, 2017). Este processo tem como objetivo identificar requisitos e/ou padrões da qualidade do projeto e suas entregas, e documentar como um projeto pode demonstrar a sua conformidade com o conjunto de requisitos e/ou padrões de qualidade pré-estabelecidos.

Planejar o gerenciamento de qualidade é a primeira etapa do processo de gerenciamento de qualidade de projetos, logo esta etapa deve se aplicar a todos os tipos de projetos, independente da sua natureza. Conseqüentemente, falhas na identificação ou documentação dos requisitos de qualidade de um projeto (ou mais especificamente, dos seus artefatos) podem ter conseqüências negativas para as partes interessadas de um projeto de pesquisa. No contexto de artefatos de projetos de pesquisa, o conjunto de partes interessadas pode incluir pesquisadores principais, profissionais da indústria, estudantes de mestrado ou

doutorado, e agências de fomento. Na Figura 1, apresentam-se os processos da atividade de gerenciamento da qualidade do projeto e suas inter-relações, com destaque ao processo de planejar o gerenciamento da qualidade.

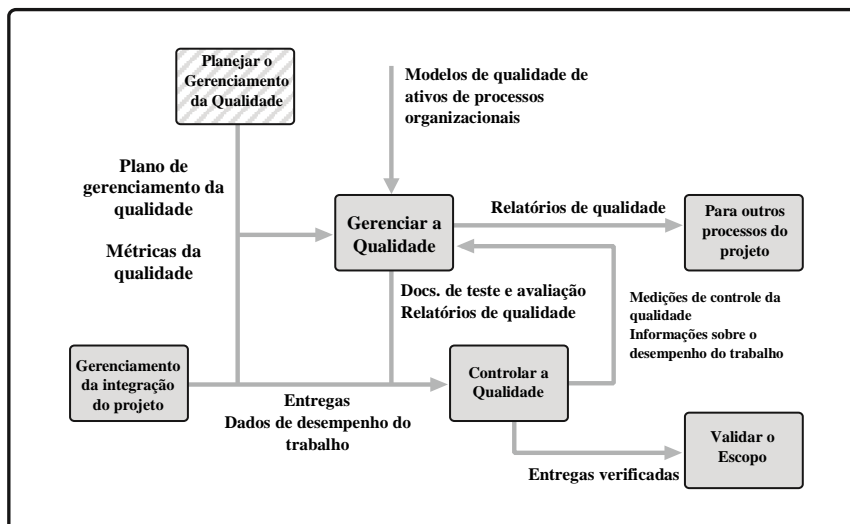


Figura 1. O processo de gerenciamento da qualidade do projeto
Fonte: PMI (2017)

No contexto deste trabalho, *requisitos de qualidade de artefatos* são entendidos como as expectativas que a comunidade de pesquisadores e profissionais da indústria com especialidade em MDE têm para com artefatos de pesquisa. Para identificar estes requisitos e padrões de qualidade, foram utilizadas as seguintes ferramentas e técnicas de gerenciamento de projetos: (i.) Análise de documentos, (ii.) Mapeamento mental, (iii.) Questionários, e (iv.) Histogramas.

Para o desenvolvimento deste catálogo, foi analisado um conjunto de nove trabalhos selecionados por meio de uma revisão da literatura de ES e MDE. Da literatura sobre de ES e artefatos de pesquisa em geral, foram analisados oito catálogos de boas práticas de pesquisa voltados para artefatos de tipos genéricos, i.e., sem foco em uma área específica. Os catálogos de boas práticas analisados foram os seguintes: (i.) “The ACM Artifact Review and Badging” (ACM, 2020); (ii.) “The EMSE Open Science Initiative” (EMSE, 2021); (iii.) “The Journal of Open Science Software” (Katz; Niemeyer; Smith, 2018); (iv.) “The Journal of Open Research Software” (JORS) (JORS, 2021); (v.) “The Guidelines by Wilson et al.” (Wilson et al., 2014); (vi.) “The NASA Open Source Software Projects” (NASA, 2021); (vii.) “The TACAS artifact evaluation guidelines” (TACAS, 2019); e (viii.) “The CAV artifact evaluation guidelines” (CAV, 2019). Da literatura específica de MDE, foram selecionados e analisados dois trabalhos: (ix.) Um artigo de revista que propõe uma taxonomia de problemas que podem afetar a adoção de ferramentas de MDE pela indústria (Whittle et al., 2017), e (ix.) um artigo de conferência que discute critérios de qualidade para repositórios de artefatos de modelagem

(Basso; Werner; Oliveira, 2017). Cada um destes documentos foi estudado a fim de catalogar lições aprendidas por especialistas do domínio e identificar lacunas nos conjuntos de práticas adotadas pela comunidade de ES e MDE.

O modelo 5W2H (também conhecido como *5W1H* ou “*Five Ws*”) é uma conhecida ferramenta usada por jornalistas para apresentar notícias e sumarizar histórias. Na perspectiva deste modelo, espera-se que repórteres sejam capazes de resumir informações sobre um fato segundo sete categorias: O quê (“What”), Onde (“Where”), Por quê (“Why”), Quem (“Who”), Quando (“When”), Como (“How”), e Quanto (“How Much”). Cada uma destas categorias constitui um tipo básico de informação que as pessoas frequentemente desejam saber sobre uma história. Além de jornalistas, este modelo também pode ser usado por pesquisadores durante o estabelecimento de perguntas de pesquisa de mapeamento sistemáticos (Jia et al., 2016) e como um “framework” de processamento de linguagem natural e sumarização de textos científicos (Prana et al., 2019; Zhang et al., 2020). Para entender o conceito de qualidade de artefato de pesquisa, o 5W2H foi proposto como modelo para categorização de práticas de pesquisa e desenvolvimento de perguntas factuais.

O mapeamento mental é uma técnica de representação de dados gráfica que se utiliza de uma estrutura hierárquica visual denominada mapa mental para apresentar conceitos e as relações entre eles (Tague, 2005). Em um mapa mental, um nó central apresenta um conceito central aonde subtópicos associados a este conceito são desmembrados na forma de nós adicionais estruturados em ramos que se irradiam a partir deste nó central. Esta é uma técnica muito útil para a representação não-linear de dados e conhecimento e em atividades de “brainstorming”. Neste trabalho, a ferramenta de mapeamento mental e o modelo 5W2H de arguição (Tague, 2005), que será discutido no próximo parágrafo, foram utilizados na categorização de boas práticas de pesquisa. Uma ilustração de um mapa mental para o 5W2H é apresentada na Figura 2.

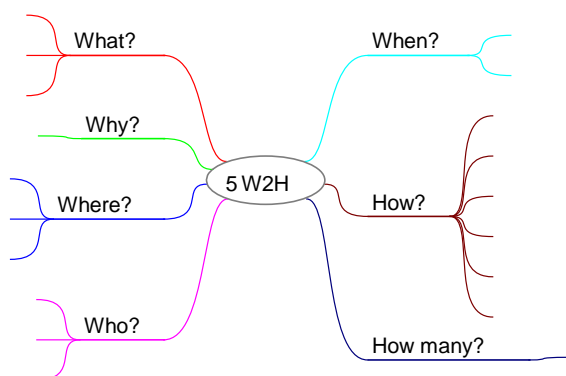


Figura 2. Exemplo de mapa mental para o modelo 5W2H
Fonte: Dados originais da pesquisa

Uma vez que um conjunto de boas práticas de pesquisa e perguntas factuais tenha sido estabelecido e categorizado segundo o modelo 5W2H, este trabalho propôs o uso de questionários e consulta a especialistas para avaliação de um conjunto de boas práticas de gerenciamento de qualidade de artefatos de pesquisas em MDE. Neste questionário, perguntas inquirindo diferentes aspectos de uma proposta de lista de boas práticas, tais como clareza, completude e relevância; foram incluídas para coletar dados quantitativos (e.g., escala Likert) e qualitativos (e.g., obtidos através de campos de texto aberto), que descrevessem a visão dos participantes quanto a lista de boas práticas de pesquisa. Uma vez coletados, dados quantitativos podem ser apresentados por meio de gráficos do tipo histograma, e.g., para auxiliar na identificação de respostas mais recorrentes.

Com base nestas ferramentas, o esquema apresentado na Figura 3 foi definido como método de estudo.

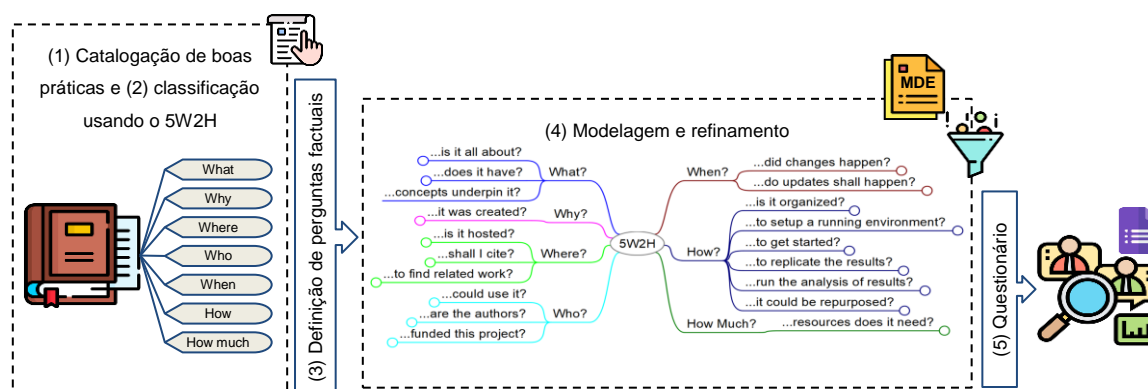


Figura 3. Método usado para desenvolvimento de boas práticas de pesquisa
Fonte: Dados originais da pesquisa

Primeiramente, propôs-se a realização de uma (1) análise documental das práticas de pesquisa com objetivo de identificar boas práticas de pesquisa genéricas. Em um segundo momento, definiu-se a (2) categorização das práticas extraídas segundo as dimensões do 5W2H. Com base nesta categorização, definiu-se o uso dos rótulos 5W2H como forma de relacionar as boas práticas de pesquisa identificadas aos diferentes aspectos que permeiam a ideia de *qualidade de artefato de pesquisa*, e.g., *saber do que se trata um artefato?*, *saber quem são seus "stakeholders"?*, *saber como usar um artefato?*. Em um terceiro momento, estabeleceu-se o uso dos rótulos 5W2H como referência para a (3) definição de perguntas factuais que pudessem guiar as partes interessadas de um artefato de pesquisa (i.e., "stakeholders" de artefatos) nos processos de criação e revisão de artefatos. Em seguida, planejou-se (4) a criação de um mapa mental que descrevesse a relação hierárquica entre as dimensões do 5W2H, as perguntas factuais, e as práticas de pesquisa (i.e., 5W2H -> rótulo

5W2H -> pergunta factual correspondente ao rótulo 5W2H -> boa prática para tratar pergunta factual).

Uma vez planejado este método, foi identificado o risco de que a lista de práticas pudesse se tornar muito extensa, e conseqüentemente ter um custo de implementação significativo ou até mesmo impraticável. A fim de tratar isso, planejou-se a realização de um refinamento deste mapa mental a fim de remover práticas redundantes. Junto disso, buscou-se também identificar lacunas relacionada a aspectos específicos da área de MDE. Práticas identificadas na literatura científica de MDE auxiliariam nesta etapa de refinamento. Por fim, propôs-se a realização de um questionário entre a comunidade de especialistas em MDE para avaliar a clareza, completude, e relevância da lista de boas práticas, segundo o ponto de vista de especialistas da área, e identificar eventuais limitações do conjunto de práticas resultante deste trabalho.

Resultados e Discussão

Com base nos oito catálogos de boas práticas, um conjunto inicial de 284 práticas foi extraído por meio de duas etapas. Primeiramente, cada um dos catálogos foi analisado a fim de identificar excertos que remetesse a boas práticas de pesquisa para compartilhamento de artefatos de tipo arbitrário. Em seguida, o conjunto de excertos extraídos foi revisado e refinado a fim de padronizar a estrutura e terminologia utilizada nas práticas. Ao término deste processo, uma lista contendo 284 boas práticas de pesquisa para artefatos arbitrários foi obtida. Exemplos de práticas extraídas e das suas respectivas versões refinadas podem ser vistos na Figura 4.

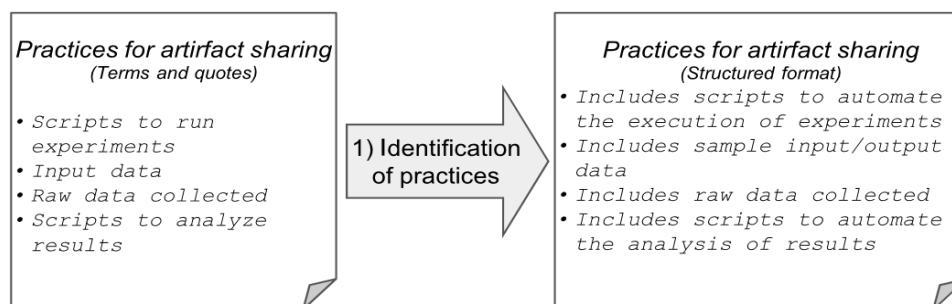


Figura 4. Exemplo de práticas extraídas e refinadas
Fonte: Resultados originais da pesquisa

Após extrair e refinar o conjunto de boas práticas de pesquisa, utilizou-se um mapa mental do modelo 5W2H para categorizar as 284 práticas enumeradas. Esta etapa partiu do pressuposto que *boas práticas de pesquisa podem ser adotadas para atender à requisitos de qualidade de artefatos discutidos na forma de perguntas factuais*. Dessa forma, o método

5W2H foi usado para obter “insights” sobre perguntas factuais que as práticas identificadas poderiam eventualmente tratar. A Tabela 1 mostra a taxa de práticas por rótulo 5W2H.

Tabela 1. Percentual de práticas por rótulo 5W2H

Rótulo	Percentual de práticas por rótulo
“What”	31,44
“Where”	17,00
“Why”	3,97
“Who”	7,93
“When”	2,83
“How”	34,84
“How Much”	1,98

Fonte: Resultados originais da pesquisa

No processo de classificação, o seguinte padrão foi usado como referência para associar rótulos 5W2H às boas práticas de pesquisa. Ao rótulo “What”, foram associadas práticas indicando a descrição do contexto de desenvolvimento e conteúdo do artefato. Ao rótulo “Where”, foram associadas práticas discutindo aspectos de hospedagem e arquivamento do artefato, citação do artefato e trabalhos relacionados. Ao rótulo “Why”, foram associadas práticas discutindo a motivação para desenvolvimento do artefato, seus objetivos e principais vantagens ou diferencial. Ao rótulo “Who”, foram associadas práticas discutindo licenças de software e direitos de uso, detalhes sobre os autores (nome, contato, afiliação) e agências de fomento. Ao rótulo “When”, foram atribuídas práticas associadas a controle de versão, identificação de versões, “roadmap” e trabalhos futuro. Ao rótulo “How”, foram atribuídas práticas associadas a configuração do ambiente experimental, replicação de experimentos, análise de resultados, e reuso de artefatos. Por fim, ao rótulo “How Many/Much”, foram atribuídas práticas associadas ao quantitativo de recursos de sistema e de tempo necessário para executar a ferramenta usando um conjunto de dados de exemplo.

Após a etapa de classificação, os sete conjuntos de práticas rotuladas foram fragmentados em 18 subconjuntos de práticas associadas. Consequentemente, para cada subconjunto, estabeleceu-se uma pergunta factual que inquirisse diferentes aspectos de qualidade relacionados a artefatos de pesquisa. Em decorrência do alto número de práticas, i.e., 284 práticas, um processo de refinamento precisou ser realizado a fim de remover eventuais redundâncias e identificar lacunas relacionadas a problemas específicos da área de MDE. Como resultado desta etapa de refinamento, o conjunto inicial de 284 práticas foi reestruturado na forma de uma lista de 77 boas práticas de pesquisas.

A fim de cobrir lacunas identificadas nesta lista de 77 práticas, consultou-se a literatura de MDE, em particular os artigos sobre problemas com ferramentas de MDE na indústria, e critérios de qualidade para repositórios de artefatos de modelagem. Com base nestes dois trabalhos, um novo conjunto de sete práticas e uma pergunta factual especificamente voltadas

para artefatos de MDE foi proposto e incluído no mapa mental. Ao término desta etapa, concluiu-se o processo de modelagem e refinamento com um total de 19 perguntas factuais e 84 práticas de pesquisas. O conjunto de 19 perguntas factuais pode ser visualizado na Figura 5. A lista de 84 práticas de pesquisa por sua vez pode ser acessada na página do projeto (MDE Artifacts, 2021), criada com o objetivo de fomentar o reuso deste trabalho.

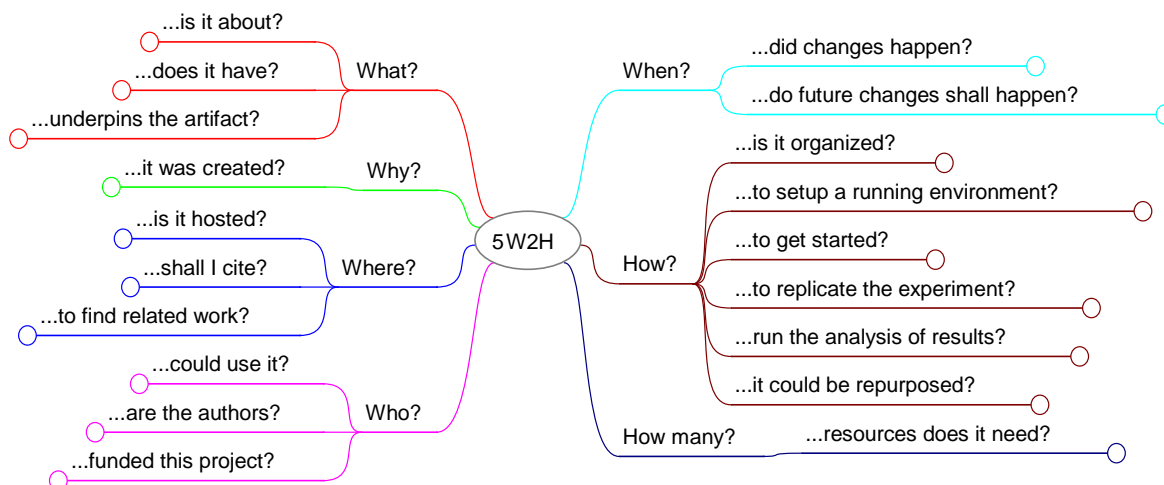


Figura 5. Mapa mental com o conjunto final de 19 perguntas factuais
Fonte: Resultados originais da pesquisa

Após o desenvolvimento deste conjunto de boas práticas, um questionário foi projetado para coletar a opinião de especialistas em MDE sobre as práticas desenvolvidas para gerenciamento de qualidade de artefatos de pesquisa em MDE. Para este questionário, foram convidados pesquisadores e profissionais da indústria com experiência em MDE. Dentre os convidados, foram incluídos ex-membros de comitês de avaliação de artefatos da “International Conference on Model Driven Engineering Languages and Systems” (MoDEL), coautores de artigos que publicados na conferência MoDELS que tenham recebido um “ACM Artifact Badge”, ou coautores de artigos publicados na revista “Software and Systems Modeling” (SoSyM) durante os três últimos anos (2020-2017) e que incluíssem algum artefato de software. Outros canais de comunicação como, lista de contatos pessoais e grupos de e-mail (e.g., PlanetMDE, SBC-L) também foram usados com o objetivo de alcançar um maior número de especialistas na área. Este questionário ficou aberto para participação durante os meses de abril e maio de 2021 e obteve uma adesão de 90 especialistas em MDE.

Para cada participante, foram apresentados seis conjuntos de perguntas cobrindo dados demográficos (i.e., gênero e cargo principal), experiência com reuso e compartilhamento de artefatos de pesquisas em MDE, dificuldades enfrentadas durante o desenvolvimento e compartilhamento de artefatos, avaliação individual das práticas (i.e., escala Likert e campo de texto aberto), e avaliação geral das práticas (i.e., escala Likert para

mensurar a clareza, relevância, completude do conjunto de práticas). Dados qualitativos sobre limitações e sugestões de melhorias também foram requisitados por meio de uma pergunta aberta. Na seção seguinte, serão discutidos os dados demográficos, alguns exemplos de práticas de pesquisa com alto nível de prioridade, a avaliação geral das práticas com relação a sua clareza, relevância e completude, e eventuais limitações e sugestões de melhoria reportadas pelos participantes. Os resultados completos deste questionário encontram-se disponíveis no site do projeto (MDE Artifacts, 2021). Adicionalmente, os artefatos gerados neste estudo também se encontram versionados e arquivados em repositórios online de compartilhamento dados popularmente usados em ciência aberta, i.e., GitHub¹ e Zenodo². Artigos complementares a este trabalho encontram-se publicados em uma conferência científica da área de MDE (Damasceno e Strüber, 2021) e em um workshop de ciência aberta (Damasceno; Melo; Strüber, 2021).

Perfil demográfico dos participantes

Com relação ao perfil demográfico dos participantes, 43.3% se auto identificou como “Academic Professor”, sendo 85.6% destes do gênero masculino. O número de acadêmicos do gênero feminino se mostrou semelhante entre os diferentes cargos, diferentemente dos acadêmicos de gênero masculino, que se apresentaram em sua maioria como “Academic Professor”. Ainda que houvesse um campo aberto para que participantes não-binários reportassem seu gênero, nenhum participante fez uso deste campo. Com relação a experiência dos candidatos, percebeu-se um alto número de participantes (i.e., 83.2%) com experiência em reuso de artefatos. Estes resultados indicaram um nível significativo de experiência coletiva entre o público de participantes deste questionário. Na Tabela 2, o número de participantes por gênero e cargo é apresentado.

Tabela 2. Dados demográficos – Principal cargo e gênero dos participantes

Gênero	Cargo Principal	Total	Percentual
Feminino	“Academic (Pre-Phd)”	4	4.4%
	“Academic (Professor)”	4	4.4%
	“Academic (Post-Doc)”	5	5.6%
	“Industrial Practitioner”	7	7.8%
	“Industrial Researcher”	7	7.8%
Masculino	“Academic (Pre-Phd)”	10	11.1%
	“Academic (Post-Doc)”	18	20.0%
	“Academic (Professor)”	35	38.9%

Fonte: Resultados originais da pesquisa

¹ Disponível em <https://github.com/damascenodiego/mdeartifacts.github.io/>

² Disponível em <https://doi.org/10.5281/zenodo.5109401>

Exemplos de práticas de pesquisa de maior prioridade

O conjunto de práticas de pesquisa apresentado aos participantes deste questionário foi estruturado em um catálogo de 84 práticas associadas a 19 perguntas factuais. Cada prática foi apresentada como uma possível forma de responder (ou tratar) as suas respectivas perguntas factuais. Cada pergunta factual, por sua vez, foi designada a partir das dimensões do método 5W2H. Para cada participante, foi solicitado que cada prática fosse categorizada como de prioridade “essencial”, “desejável” ou “desnecessário”. A Tabela 3 apresenta o subconjunto de 23 práticas de pesquisa com prioridade categorizada como essencial por pelo menos 50% dos participantes. O percentual de votos atribuído a cada prioridade está ilustrado de segundo a intensidade de cor cinza em cada célula.

Tabela 3. Lista de práticas para compartilhamento de artefatos de alta prioridade (Top 23)

5W2H	Pergunta	Prática	Prioridade (% de votos)		
			Essenc.	Desej.	Desn.
What	1.1) What is it all about?	Indicate the context of development	64,40%	33,30%	2,20%
		Report artifact name	71,90%	24,70%	3,40%
		Indicate main functionalities supported	81,10%	18,90%	0,00%
	1.2) What does it have?	Include everything required for replications	78,90%	21,10%	0,00%
		1.3) What underpins the artifact?	Indicate modeling languages used to develop the artifact (e.g., UML, SysML,)	55,10%	40,40%
	Indicate libraries/frameworks used and their respective versions (e.g., Eclipse release)		62,20%	33,30%	4,40%
Why	2.1) Why it was created?	Indicate objective (e.g., replicability, reusability)	51,10%	44,40%	4,40%
Where	3.1) Where is it hosted?	Repository is open and public (e.g., GitHub, Zenodo, Figshare)	77,80%	17,80%	4,40%
	3.3) Where to find related work?	Give credit to data obtained from other sources (e.g., author, repository)	58,90%	41,10%	0,00%
Who	4.1) Who could use it?	Deposited under explicit open license (e.g., reported in a LICENSE file)	53,90%	41,60%	4,50%
	4.2) Who are the authors?	Indicate the authors names	73,30%	24,40%	2,20%
	4.2) Who are the authors?	Indicate the authors's contact details (e.g., email, ResearchGate, website)	56,70%	36,70%	6,70%
When	5.1) When did changes happen?	Tracked using version control (e.g., GitHub, GitLab, BitBucket)	61,10%	36,70%	2,20%
How	6.1) How is it organized?	Files and folders shall have self-explaining names matching content	51,70%	47,20%	1,10%
		6.2) How to setup a running environment?	Has step-by-step info to build the code	60,70%	37,10%
	Provides instructions for downloading		77,50%	18,00%	4,50%
	Provides instructions to install it		86,70%	12,20%	1,10%
	6.3) How to get started?	Includes minimal test data	62,90%	33,70%	3,40%
		Include step-by-step instructions to run it	86,70%	13,30%	0,00%
	6.4) How to replicate the experiment?	Provide manual/automated instructions for the complete/partial replications	57,30%	41,60%	1,10%
Includes the complete set of test models		58,40%	39,30%	2,20%	
6.5) How to run analysis of results?	Provides a clear description of measurements/metrics used	51,70%	41,60%	6,70%	
How Many	7.1) How many resources are needed?	Indicates system/environment settings where it was successfully evaluated	53,30%	42,20%	4,40%

Fonte: Resultados originais da pesquisa

Avaliação geral do conjunto de práticas

Para avaliar a opinião dos especialistas com relação a completude, clareza e relevância da lista de práticas proposta, foi solicitado que cada participante reportasse uma pontuação numa escala Likert de sete pontos. Na Figura 6, a frequência de pontuações atribuídas a clareza, completude e relevância do conjunto de boas práticas são apresentadas em histogramas com suas respectivas medianas indicadas por linhas pontilhadas.

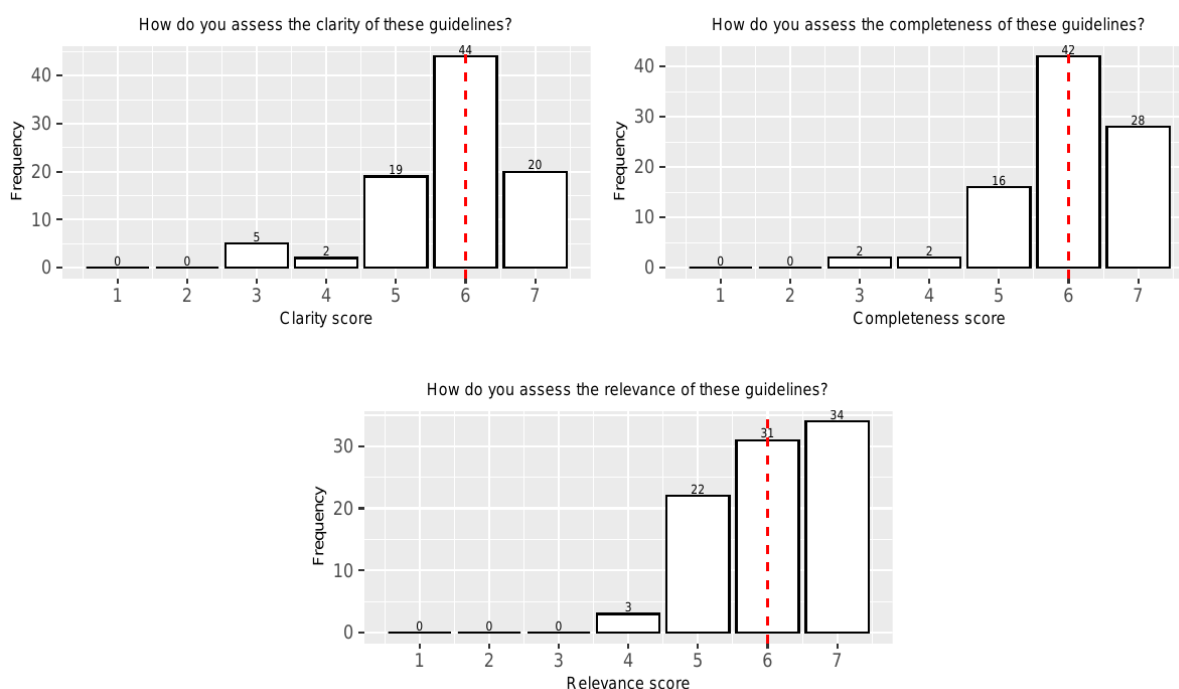


Figura 6. Histogramas dos níveis de clareza, completude e relevância das práticas reportados pelos especialistas em MDE
Fonte: Resultados originais da pesquisa

Em geral, mais de 92% dos participantes reportaram pontuações altas para as três perspectivas avaliadas, i.e., entre 5-7 na escala Likert. Para a completude, clareza e relevância, foi identificado que 95.5%, 92.2%, e 96.6% dos participantes reportaram pontuações positivas. Nenhum participante reportou pontuação abaixo dos três pontos.

Comentários, limitações, e sugestões de melhoria

Dentre os comentários obtidos no campo de texto, foram identificados vários relatos positivos que corroboram as pontuações positivas reportadas, dentre eles: “A ampla implementação de diretrizes como essas é *ESSENCIAL* para o avanço das tecnologias e pesquisa de MDE!” (Tradução livre). Já com base no “feedback” textual dos participantes que

reportaram pontuações baixas, foram identificados elementos a serem melhorados na lista de boas práticas de pesquisa, tais como a falta de foco em artefatos recentemente criados, e os vários significados que a palavra compartilhamento pode ter dependendo do seu contexto ou “stakeholder”. Em particular, no contexto de pesquisa na indústria, o compartilhamento de artefatos é mencionado como uma tarefa desafiadora em decorrência do caráter proprietário e confidencial que certos ativos podem adquirir.

Conclusão

Este trabalho apresentou um processo para auxiliar no desenvolvimento e validação de um conjunto de boas práticas de pesquisa para gerenciamento de qualidade de artefatos de pesquisa científica em MDE. Artefatos de pesquisa vêm sendo cada vez mais valorizados pelo seu potencial de auxiliar a reprodução de estudos científicos, desenvolvimento de novas tecnologias e adoção de abordagens inovadoras pela indústria. A fim de fomentar uma melhoria na qualidade dos artefatos de pesquisa em MDE, um conjunto de princípios de gerenciamento de qualidade de projetos foi aplicado para catalogar e refinar um conjunto de boas práticas de pesquisa. Este conjunto de boas práticas foi avaliado por meio de um questionário realizado entre especialistas da área. A partir deste questionário, especialistas em MDE, ocupando diferentes cargo e com significativa experiência no uso e compartilhamento de artefato, reportaram altos níveis de satisfação, no que tange a completude, relevância e clareza da lista de práticas proposta. Como trabalhos futuros, pretende-se realizar um estudo utilizando técnicas de priorização multicritério para identificar quais das práticas e perguntas factuais discutidas devem ser consideradas como de maior prioridade. Dessa forma, pesquisadores de MDE poderão adaptar as práticas apresentadas neste estudo de acordo com suas necessidades e demandas de projeto mais emergenciais.

Referências

Association for Computing Machinery [ACM]. Artifact Review and Badging - Current. Disponível em: <<https://www.acm.org/publications/policies/artifact-review-and-badging-current>>. Acesso em: 18 mar. 2021.

Basciani, F.; Di Rocco, J.; Di Ruscio, D.; Iovino, L.; Pierantonio, A. 2019. A tool-supported approach for assessing the quality of modeling artifacts. *Journal of Computer Languages* 51: 173–192.

Basso, F. P.; Werner, C. M. L.; Oliveira, T. C. Revisiting Criteria for Description of MDE Artifacts. 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)*. 11: 27-33.

Damasceno, C. D. N.; Melo, I. C.; Strüber, D. 2021. Towards Multi-Criteria Prioritization of Best Practices in Research Artifact Sharing. Anais do Workshop de Práticas de Ciência Aberta para Engenharia de Software (OpenSciense). 1: 1-6

Damasceno, C. D. N.; Strüber, D. 2021. Quality Guidelines for Research Artifacts in Model-Driven Engineering. ACM/IEEE 24th International Conference on Model Driven Engineering Languages and Systems (MODELS 2021). Disponível em: <<http://arxiv.org/abs/2108.04652>>. Acesso em: 25 ago. 2021

Hermann, B.; Winter, S.; Siegmund, J. 2020. Community expectations for research artifacts and evaluation processes. Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. 28: 469-480.

Heumüller, R.; Nielebock, S.; Krüger, J.; Ortmeier, F. 2020. Publish or perish, but do not forget your software artifacts. Empirical Software Engineering 25: 4585–4616.

The MDE Artifacts Project [MDE Artifacts]. 2021. Disponível em: <<https://mdeartifacts.github.io/>>. Acesso em: 15 jul. 2021.

International Conference on Computer-Aided Verification [CAV]. 2019. Artifacts | CAV 2019. Disponível em: <<http://i-cav.org/2019/artifacts/>>. Acesso em: 18 mar. 2021

International Conference on Tools and Algorithms for the Construction and Analysis of Systems [TACAS]. 2019. TACAS - ETAPS 2019. Disponível em: <<https://conf.researchr.org/track/etaps-2019/tacas-2019-papers#Artifact-Evaluation>>. Acesso em: 18 mar. 2021.

Jia, C.; Cai, Y.; Yu, Y. T.; Tse, T. H. 2016. 5W+1H pattern: A perspective of systematic mapping studies and a case study on cloud software testing. Journal of Systems and Software 116: 206–219.

Journal of Empirical Software Engineering [EMSE]. 2021. EMSE Open Science Initiative: Principles, Board, Process and FAQ. Disponível em: <<https://github.com/emsejournal/openscience/blob/master/README.md>>. Acesso em: 18 mar. 2021.

Katz, D. S.; Niemeyer, K. E.; Smith, A. M. 2018. Publish your software: Introducing the Journal of Open Source Software (JOSS). Computing in Science Engineering 20: 84–88.

Marjan Grootveld; Ellen Leenarts; Sarah Jones; Emilie Hermans; Eliane Fankhauser. 2018. OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans. Disponível em: <<https://zenodo.org/record/1120245>>. Acesso em: 5 abr. 2021.

Penzenstadler, B.; Femmer, H. 2013. A generic model for sustainability with process- and product-specific instances. Proceedings of the 2013 workshop on Green in/by software engineering. Acesso em: 5 abr. 2021.

Prana, G. A. A.; Treude, C.; Thung, F.; Atapattu, T.; Lo, D. 2019. Categorizing the Content of GitHub README Files. Empirical Software Engineering 24: 1296–1327.

Project Management Institute [PMI]. 2017. A guide to the project management body of knowledge / Project Management Institute. 6ed. Project Management Institute, Newtown Square, PA, EUA.

Tague, N. R. 2005. The quality toolbox. 2ed. ASQ Quality Press. Milwaukee, Wis, EUA.

The Journal of Open Research Software [JORS]. 2021. The JORS - Editorial Policies. Disponível em: <<http://openresearchsoftware.metajnl.com/about/editorialpolicies/>>. Acesso em: 18 mar. 2021.

The National Aeronautics and Space Administration [NASA]. 2021. NASA Open Source Software. Disponível em: <<https://code.nasa.gov/>>. Acesso em: 18 mar. 2021.

Van den Eynden, V.; Corti, L.; Woollard, M.; Bishop, L.; Horton, L. 2011. Managing and Sharing Data; a best practice guide for researchers. Disponível em: <<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>>. Acesso em: 17 mar. 2021.

Whittle, J.; Hutchinson, J.; Rouncefield, M.; Burden, H.; Heldal, R. 2017. A taxonomy of tool-related issues affecting the adoption of model-driven engineering. *Software & Systems Modeling* 16: 313–331.

Wilson, G.; Aruliah, D. A.; Brown, C. T.; Hong, N. P. C.; Davis, M.; Guy, R. T.; Haddock, S. H. D.; Huff, K. D.; Mitchell, I. M.; Plumbley, M. D.; Waugh, B.; White, E. P.; Wilson, P. 2014. Best Practices for Scientific Computing. *PLOS Biology* 12: 1-7.

Zhang, J.; Li, K.; Yao, C.; Sun, Y. 2020. Event-based summarization method for scientific literature. *Personal and Ubiquitous Computing*.