



Towards Open Science at the DELFI Conference

Natalie Kiesler ¹, René Röpke ², Daniel Schiffner ³, Sandra Schulz ⁴, Sven Strickroth ⁵, Matthias Ehlenz ⁶, Birte Heinemann ⁷, and Arno Wilhelm-Weidner ⁸

Abstract: Despite the increasing awareness of Open Science within the educational technology community, conferences, such as DELFI, do not yet foster the publication of research data including software. To address this, we conducted a survey eliciting the community's needs, perspectives, and publication preferences. The analysis of 24 valid responses reveals a variety of research data formats used, and several uncertainties, e. g., regarding data ownership. Associated barriers comprise legal concerns and lacking resources to publish data. Nonetheless, researchers seem open for new publication formats. Moreover, we analyzed author's intentions to publish data related to their DELFI submissions in 2023 (n=66). Many researchers assume not to have data to share (n=28), or no intention to publish data in the future (n=16). Overall, the results imply a lack of awareness and recognition of data publications, so that further efforts and incentives are required to move toward Open Science practices in the DELFI community.

Keywords: Open Science, Open Data, Data publication, FAIR, DELFI, Educational Technology

1 Introduction

In educational technology research (EdTec), as in most scientific domains, the push towards Open Science has gained significant momentum, underscored by the increasing demands of funding institutions for greater transparency and accessibility. Initiatives such as the Open Science Framework [FD17] and Open Data mandates by major organizations [Eu23; Na13; PF07] highlight a global shift towards open research practices that are more inclusive, collaborative, and accessible to a broader audience. Findability, Accessibility, Interoperability, and Reusability [Wi16] have emerged as core principles (FAIR), providing a comprehensive guideline for enhancing the openness and usability of scientific data.

- 1 Nuremberg Tech, Computer Science, Keßlerplatz 12, 90489 Nuremberg, Germany, natalie.kiesler@th-nuernberg.de, <https://orcid.org/0000-0002-6843-2729>
- 2 TU Wien, Learning Technologies and eDidactics, Favoritenstr. 9-11, 1040 Vienna, Austria, rene.roepke@tuwien.ac.at, <https://orcid.org/0000-0003-0250-8521>
- 3 DIPF | Leibniz Institute for Research and Information in Education, Rostocker Str. 6, 60323 Frankfurt am Main, Germany, d.schiffner@dipf.de, <https://orcid.org/0000-0002-0794-0359>
- 4 University of Hamburg, Faculty of Education, Von-Melle-Park 8, 20146 Hamburg, Germany, sandra.schulz@uni-hamburg.de, <https://orcid.org/0000-0002-2254-6579>
- 5 LMU Munich, Institute for Informatics, Oettingenstraße 67, 80538 Munich, Germany, sven.strickroth@ifi.lmu.de, <https://orcid.org/0000-0002-9647-300X>
- 6 RWTH Aachen University, Lehrerbildungszentrum, MediaLab Lehramt, Kármánstr. 17-19, 52056 Aachen, Germany, ehlenz@lbz.rwth-aachen.de, <https://orcid.org/0000-0001-6189-6056>
- 7 RWTH Aachen University, Learning Technologies, Ahornstr. 55, 52074 Aachen, Germany, heinemann@cs.rwth-aachen.de, <https://orcid.org/0000-0002-7568-0704>
- 8 VDI/VDE-IT, Bildung und Wissenschaft, Steinplatz 1, 10623 Berlin, Germany, arno.wilhelm-weidner@vdivde-it.de, <https://orcid.org/0000-0003-2604-3327>

However, it is not common practice in the EdTec community to publish research data (including software) along with paper publications [KS22; KS23a]. Especially the DELFI conference by the German Informatics Society (GI) offers little room for the publication of research data, as its submission is not obligatory [Ge24b]. The limited availability of research data from DELFI papers published between 2018 and 2022 [MS23] confirms the lack of Open Data practices. We, as a community, thus need to critically assess and improve our practices to foster an alignment with Open Science.

This paper aims to provide insights into the current state of Open Science practices within the EdTec community, specifically focusing on the DELFI as a German EdTec conference. Based on the results of an online survey, we examine the community's perspective on publication practices, the characteristics of the research data, the barriers and challenges to open publication, and the preferences and expectations of researchers. Furthermore, we explore the intentions of authors with papers accepted at DELFI 2023 regarding the publication of their data, leveraging information gathered via the paper submission system. By providing these insights, we aim to start drafting a path forward towards Open Data practices at future DELFI conferences.

2 Background

Finding, accessing, understanding, and reusing data produced by others are challenges reported by many researchers. Likewise, replicating prior work is often difficult due to the lack of sufficient details [Ih15; Je22; Ki23; Pr23]. Suppose software is used as a method for research. In that case, it is hardly recognized within the community, rarely published, and thus not available, causing the reinvention of the wheel over and over again [KS22]. In particular, we observe the following challenges related to searching and reusing data: (1) Lack of (long-term) availability, e. g., loss of files, limitations through proprietary or deprecated software [GNT10; KGH21; KPC21]; (2) Lack of comprehensibility, and data maturity e. g., missing documentation due to required effort, time, and resources; (3) Lack of recognition for high quality, mature research data, e. g., publication is not recognized within the academic community [Mo11]; (4) Lack of obligatory reviews and guidelines, i. e., data submissions are rarely allowed/supported, and data review processes are not defined [KS23b]; and (5) Lack of data provenance, the data's origin, data preprocessing and cleaning steps, errors, or other aspects relevant for secondary research [CF10].

At the same time, publishing data can entail several obstacles for researchers. For example, Strickroth and Striewe [SS22] had to publish their research data on a personal website due to missing features on existing platforms and the lack of workflows for collaborative maintenance. Other limitations apply to research data centers, such as the German Centre for Higher Education Research and Science Studies [Ge23a]. Data publication processes are demanding and involve several time-consuming steps (over several months), people, and additional metadata and reports [Ki22a; Ki22b; SBH23]. Irwin [Ir13] adds ethical concerns due to the sensitivity of some (qualitative) data.

In an interview study with 13 researchers (from Germany, Peru, India, and China) in the context of learning analytics, a number of barriers preventing researchers from the publication of their data were identified. Among them are the following categories [BP21]:

- Legal concerns – barriers related to copyright or other legal considerations that may impact data publication;
- Fear of loosing control – barriers concerning the sovereignty over the data;
- Authority or practice considerations – established guidelines or research practices within a particular field affecting the acceptance of data publications;
- Technical / processing barriers – challenges associated with the publication process;
- Resource barriers – e. g., limited funding, human resources, or infrastructure hindering data publications.

Even though these concerns can vary across countries, the human and financial resources required for the preparation and publication of data, and the lack of infrastructure remain a common barrier, especially in countries with low incomes. Another result of the study is that most of the researchers have not published their data [BP21].

In a conservative estimate, a report on behalf of the European Commission calculates opportunity costs of 10.2 billion Euros due to the lack of FAIR research data [CRI18]. McGill [Mc19] concludes the lack of open research practices contributes to decreasing research standards. Therefore, it is crucial for EdTec researchers and the DELFI community to start focusing on research data and to foster data publications, including software.

3 Methodology

This paper aims to gather the EdTec community's perspective on the publication of research data so that we can start developing feasible formats for researchers to publish their data at the DELFI conference. Our research questions are as follows:

1. How can we characterize research data in the EdTec context?
2. What are the barriers and challenges associated with the publication of research data in the EdTec context?
3. What are the expectations and preferences regarding the publication formats of research data at the DELFI conference?
4. What were the author's intentions regarding the publication of their research data at the DELFI conference in 2023?

To address RQ1, RQ2, and RQ3, we adapted an available survey instrument [Bi23], which is aligned with the 2016 Open Data Survey framework [Ce17]. Moreover, we tailored it to the specifics of the DELFI community concerning the definition of research data formats and publication formats. The survey also incorporated common barriers for researchers attempting to publish their data. These barriers are categorized into five areas: (1) legal concerns, (2) loss of control of data, (3) technical constraints, (4) authority or practice

considerations, and (5) resource constraints [BP21]. In addition, the survey included questions about potential conference formats and how they accommodate the requirements for publishing research data, including software (see Appendix A). Survey questions Q1–Q8 address RQ1, questions Q9–Q14 address RQ2, and survey questions Q15–Q18 help answering RQ3. Most questions are multiple-choice questions, with few open input answering options (see also [Bi23]).

As part of the survey introduction, respondents were provided with a definition of research data based on the German Research Foundation (DFG), which includes, among others, “*measurement data, laboratory values, audiovisual information, texts, survey or observation data, methodological test procedures, and questionnaires. Compilations and simulations can likewise constitute a key outcome of academic research (. . .). The same applies if the software is required to create or process research data.* [Ge23b]”. At the beginning of the survey, we asked participants to think of a recent representative educational research project in which research data had been created or gathered. The full survey will be available [upon acceptance]. Conducting the survey was approved by the Ethics Board of the DIPF | Leibniz Institute for Research and Information in Education.

The online survey was distributed between September and December 2023 within the German-speaking EdTec community following the purposeful sampling method [Pa02]. It should be noted that this community is considerably small. Specifically, we distributed our survey at the DELFI conference 2023, the GI’s Doctoral colloquium workshop, and via the GI’s educational technology group newsletter. This led to 58 responses (18 full, 40 partial). Due to the high number of partial responses, we had to define inclusion/exclusion criteria: (1) The processing time of the survey should be longer than 120 seconds, and (2) respondents had to fill out at least the initial five questions to remain in the sample. Responses that did not meet these requirements were ignored from the analysis. To analyze the remaining 24 response sequences (18 full, 6 partial), we use frequency and descriptive statistics for the quantitative results and thematic analysis [CB14] for open-ended, qualitative responses. Representative quotes highlight and contextualize the identified themes.

To determine the authors’ intentions to publish their research data in the DELFI context (RQ4), an additional evaluation was conducted as part of the DELFI 2023 paper submission process. All authors were asked to specify their intentions by selecting one of the following options: (a) *I have already published my research data*, (b) *I will publish my research data*, (c) *I do not want to publish any research data* and (d) *I do not have any research data for publication*. Authors were also asked to provide links or references to their data. For the analysis, only the authors’ responses from the 66 accepted papers (in the categories of long paper, short paper, demo, or poster) were considered. Moreover, if authors indicated the intention to publish their data or to have published their data already, we reviewed the DELFI 2023 proceedings, i. e., to examine whether links to external resources and data were provided.

4 Results

In this section, we summarize the findings for each research question, i. e., data characteristics (RQ1), associated barriers (RQ2), researchers' expectations and preferences (RQ3), and their intentions to publish data (RQ4). Illustrative figures highlight some of the results. The research data is published in an online repository [Ki24b].

4.1 Characterization of the Research Data (RQ1)

To address the first research question, we asked a number of questions related to the research data commonly gathered or used in EdTec projects. The goal of the respective survey questions was to characterize the research data (Q1–Q5), to identify data ownership and currently used publication formats (Q6–Q8).

First, the participants were asked what research data they used in one of their last EdTec projects (see Appendix A). Most of them reported using software developed by themselves ($n_t=16$) or by others ($n_o=16$), qualitative data collected by themselves ($n_t=18$) or by others ($n_o=8$), and quantitative data collected by themselves ($n_t=17$) or by others ($n_o=7$). For the answer option *other data*, the video and audio formats were selected by two participants.

In response to Q2, the following specific research formats were selected: plain text ($n=16$), structured text ($n=18$), general purpose formats ($n=19$), domain-specific formats ($n=7$), multimedia ($n=17$), compiled binary artifacts ($n=8$), and source code ($n=14$). As part of Q3, the participants chose the following research formats (see Fig. 1), which are important to be made available to others for the purpose of validation: plain text ($n=12$), structured text ($n=16$), general purpose formats ($n=14$), domain-specific formats ($n=4$), multimedia ($n=13$), compiled binary artifacts ($n=5$), and source code ($n=11$). Additionally, some participants indicated that none of these research data formats are important to be made available ($n=2$).

Regarding their last EdTec project, the participants described the volume of data they created and intended to share (Q4). Six participants reported having data between 10 and 400 megabytes. Other participants described a volume of 3 to 20 gigabytes ($n=11$). No one reported to have a volume of terabytes. Only one person did not know the volume of their

Research data formats to be shared

Plain text: e.g., unstructured text, essays, etc.

Structured text: e.g., XML, JSON, YAML, etc.

(Proprietary) general purpose formats: e.g., Office products (...)

(Proprietary) domain-specific formats: e.g., SCORM, Sensor data

Multimedia: e.g., Image, Video, Audio, etc.

Compiled binary artifacts: e.g., Models, Executables

Source Code: e.g., in Java, C

None of these

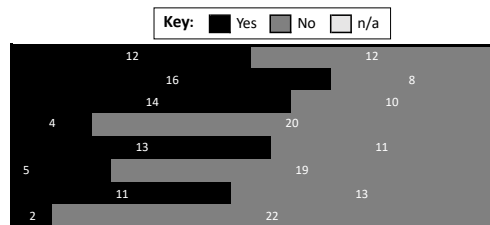


Fig. 1: Reported responses regarding research data formats to be shared (Q3)

data. Furthermore, researchers were asked for the approximate number of data files created in the last EdTec research project (Q5). Some participants did not know (n=8), and others (n=13) chose values between 5 and 5000.

When asked about data ownership (Q6), three respondents answered that they did not know who owned the data before or after publication. Most respondents answered that the data belonged to either *themselves* (n=12), *the project* (n=8) or *the institution* (n=5) before publication. After publication, it was assumed that the data belongs to *them* (n=8), *the project* (n=7), *the institution* (n=5) or others (n=5).

In response to Q7, most researchers stated they had previously published research data either in a *public software repository*, e. g., *GitHub*, *GitLab*, *DockerHub* (n=12), in a *publication's appendix* (n=7), in a *repository of their institution* (n=6) or as a *pure data publication* (n=5). The used publication platforms were *OSF* [Ce24], *Zenodo* [CE24], and *ArXiv* [Co24].

There were only a few responses to Q8, which asked for reasons why research data was not published. The most common responses are *too many barriers to overcome* (n=4), that *the research data was not sufficiently documented* (n=3), or that *the process of anonymization was not clear* (n=3). In the free text responses, it was noted that not all data could be anonymized, that the data was required for further processing, or that the study subjects had not given consent.

4.2 Associated Barriers (RQ2)

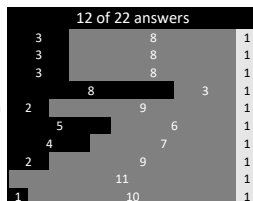
The second research question aims to identify barriers and challenges associated with the publication of research data in the EdTec context. In the survey, Q9–Q14 address well-known barriers, and specific challenges.

The barriers were prompted in a two-layer approach. First, participants could select up to five superordinate barriers (Q9) they faced in their last EdTec research project. Second, for each selected barrier area, subordinate, more specific barriers and concerns were surveyed (Q10–Q14). The respondents received those survey questions depending on their selection of barriers in Q9. Item Q9 was answered by 22 participants. Five of the participants did not experience any of the given superordinate barriers in their last research project. The others had experienced between one and five barriers (median: 2). The most frequently experienced barrier in the last project relates to *legal concerns* (n=12), followed by *resource constraints* (n=10), fears for *loss of control* (n=9), *technical/processing constraints* (n=5, e. g., data set too big, complicated), and *authority or practice considerations* (n=4).

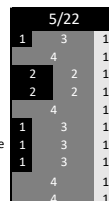
In the following, interesting results of Q10–14 are highlighted (see Figure 2). Two participants discontinued the questionnaire after Q9, resulting in *n/a* answers for Q10–Q14. Participants experienced up to seven legal barriers or concerns (median: 2, Q10). The most frequently selected option in Q10 (legal concerns) was *It is difficult to balance privacy and openness* (n=8), followed by *I'm not sure how to deal with personal or sensitive data* (n=5), and

Legal concerns (Q10)

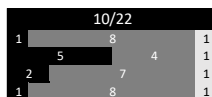
I'm unfamiliar with the legal regulations
 I'm not sure who owns the research data
 I'm not allowed to publish the research data
 It is difficult to balance privacy and openness
 Unclear who will be accountable for the (...) data
 I'm not sure how to deal with (...) sensitive data
 Anonymization process is not sufficiently...
 I'm not sure which license to choose
 The research data is subject to patents...
 I can't say

**Technical constraints (Q13)**

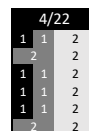
I'm unfamiliar with the publication process
 I require an embargo period for publication
 Data files are too large
 Data is too complex
 Overall publication process is too complex
 Unclear where I should publish my research data
 Initial anonymization process is not comprehensive
 Data curation (e.g. metadata) is too complex.
 Service (...) I published in doesn't exist anymore
 I can't say

**Resource constraints (Q14)**

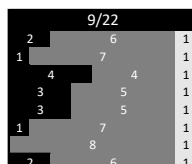
Publication of research data is too expensive
 (...) requires too much time and/or work effort
 I did not find a suitable service (...) to publish my data
 I can't say

**Authority or practice considerations (Q12)**

Data is not peer-reviewed
 Analysis is easier to understand than the raw data
 I am under no obligation to publish research data
 Little or no value for me as a researcher (...)
 Publishing (...) data not common in my discipline
 I can't say

**Loss of control of data (Q11)**

I'm not confident in the quality of my research data
 Anonymization leads to a loss of information...
 I'm afraid that somebody could misinterpret my (...) data
 Publishing data exposes any errors I made
 Somebody could answer (...) questions before me
 Someone may alter my data and republish it
 The data is viewed/(...) for commercial advantage
 I can't say



Key: Yes No n/a

Fig. 2: Overview of the selected barriers and concerns. The number of researchers with issues in an area (based on Q9) is shown in each table header. The blocks below show the responses to Q10–Q14.

Anonymization process is not sufficiently secure (n=4). Three researchers explicitly selected *I'm not allowed to publish the data*. Nobody experienced problems with patents. Up to five barriers or concerns concerned the *Loss of control* (Q11) (median: 0). Two participants selected *no* for all options, and two selected *I can't say* as well as *no* for all other options. The most frequent reply was *I'm afraid that somebody could misinterpret my research data* (n=4). The following options were selected by three participants each: *somebody could misinterpret my research data*, and *Somebody could answer one of my research questions before me*. There were no reported concerns about research data being commercially exploited. All researchers (n=2) answering Q12 about *authority or practice considerations* selected *no* or *I can't say*, or *Analysis is easier to understand than the raw data*. The remaining replies are inconclusive (see Figure 2). Two of the five researchers who experienced barriers and concerns regarding *technical/processing constraints* did not select any of the provided options in Q13. Up to five barriers and concerns were experienced by the other three researchers (median: 0). The most frequently selected options were *Data files are too large* and *Data is too complex*, both selected twice. Four respondents each explicitly stated that the following aspects are not perceived as barriers: *I require an embargo period for publication*, and *Overall publication process is too complex*. Regarding the *resource constraint* (Q14), researchers often stated that the *Publication (...) requires too much time and/or work effort* (n=5). Two researchers selected *I did not find a suitable service* and according to one researcher, *Publication (...) is too expensive*. One researcher selected *I can't say*.

In the additional open-ended answer fields of Q10–Q14, the anonymization of audio recordings was mentioned as an issue. Furthermore, two respondents stated each that the additional effort is not recognized by the community: “on the contrary, you get punished if

you try”. In addition, data tracks were identified to be missing, and the competencies of reviewers were questioned.

4.3 Expectations and Preferences for Data Publications at DELFI (RQ3)

Regarding EdTec researchers’ expectations, another series of questions were asked, including past experiences with publishing, data formats and data uploads (Q15–Q17). Another question addressed how reviewers should handle these submissions (Q18).

When asked about the publication of the research project (not the data), 15 out of 20 respondents said that their research is or will be published in the context of an EdTec conference (Q15). Only 4 could not answer this question and chose *I can’t say*. In the additional input field, two respondents added that the data could not be published or the project was not considered to be public.

Q16 addressed current conference formats and their appropriateness, and which new formats should be considered. We provided two options on existing formats (*Demo/Tools* and *Full paper tracks*) that could be extended with specifics of research software and data. Of 19 responses, 13 agreed to a more detailed *Demo/Tools track*, including additional materials and guidelines. Eleven selected the *Full paper track* and support the addition of data material and software. For the new tracks, a slight majority chose a six-page format for software within a new *Tools track* (11 out of 19) compared to the twelve-page variant. 13 participants selected a specialized *Data paper track* only for research data comprising six pages. In general, there is a positive trend towards the inclusion or extension of existing tracks with research software and research data (59%).

When asked for supplementary material and the upload of metadata and data (Q17) along with a publication, ten participants out of 15 responded with yes, confirming its necessity. This includes the possibility of providing hyperlinks to external resources.

The last question (Q18) focused on researchers’ expectations when reviewing software and research data. 14 participants answered the question and provided several common themes in their responses: Reviewers require guidelines how the artifacts should be evaluated, with FAIRness and reproducibility as driving factors. Quality is also considered important, but only on the metadata level. The participants included examples to illustrate what might help guide or incentivize the authors of research data, e. g., tutorials to setup the software, explanations how to run/read the files, or Badges.

4.4 Authors’ Intentions to Publish Research Data at DELFI (RQ4)

The analysis of authors’ intentions to publish their data reveals that for 28 of the 66 papers (42%), authors stated to have no research data for publication. For 19 of the 66 papers, the

authors have no intention to publish their research data. Moreover, only 3 authors indicated they had published their research data already. Indeed, all of these three published papers contain hyperlinks to respective repositories. Two point towards source code, while one link points towards Zenodo [CE24].

Among the 16 (out of 66) papers for which authors indicated the intention to publish their research data, there are only 4 published papers with a link or reference to a data repository. These links point to code repositories (e. g., GitLab or GitHub). For the other 12 of these 16 papers, the authors did not seem to have published their data within the scope of the DELFI 2023 proceedings. However, they may or may not have published the data elsewhere.

In Table 1, we summarize the authors' intentions by paper type. Taking a look at the 6 full papers, which by the community's definition [Fa24] require the presentation of advanced or completed research work and therefore may contain empirical data or developed software, authors indicated the intention to publish the research data in only one of the six cases. The published paper, however, contains no reference or link to published data. For the five remaining full papers, the authors did not state the intention to publish any data (4) or do not have any data for publication (1). Similarly, for the 21 demo papers, most authors indicated they have either no data for publication (9) or do not want to publish it (5). For the remaining seven demo papers with authors' intentions to publish the data, only three papers contained links to published research data.

	full	short	demo	poster	Σ
No data for publication	1	8	9	10	28 (42 %)
No intention to publish	4	6	5	4	19 (29 %)
Intention to publish in the future	1	6	6	3	16 (24 %)
Already published the data	0	2	1	0	3 (5 %)

Tab. 1: Authors' intentions to publish the research data classified by paper type.

5 Discussion

The survey results reveal several points worth discussing with regard to our research questions. For brevity, we highlight and discuss the most interesting findings before pointing out the limitations of this work.

Regarding RQ1, survey results show that there is a high heterogeneity of data sources, types, and usages. One observation is the potential misunderstanding of Q1, where respondents may have had varying interpretations of the "data by others" answer option. It was selected by 16 respondents, which was an unexpectedly high number. Respondents may have referred to coauthors or doctoral students when choosing this option. Concerning data ownership (Q6), many respondents seem uncertain about who owns the data before and after publication. Results indicate varying assumptions, with some attributing ownership to themselves,

the project, or the institution. The term “the project” raises questions about its definition and seems to indicate false assumptions within the research community. While software repositories are widely used for sharing research outputs, the survey suggests that they may not be universally suitable for different types of data (Q7). This finding highlights the need for diverse and specialized data repositories tailored to accommodate the variety of data formats and research domains.

In response to RQ2, missing recognition and large efforts to prepare data remain key challenges. Interestingly, the question addressing barriers to data sharing itself emerged as a potential barrier, as indicated by a notable drop in the number of responses. Researchers may find the topic challenging, they might have been overoptimistic in their claimed data publication goal in the beginning, or barriers were not well-defined. Particularly, legal concerns on balancing privacy and openness are perceived to limit data publication practices. These findings underline the need for targeted efforts to reduce these barriers and streamline the data sharing process.

Looking at the expectations for data publication (RQ3), we see that the author’s willingness to publish software or research data seems rather high in principle. However, the survey results show a demand of more appropriate formats that actually reflect the challenges related to creating software or processing research data. Especially defining individual tracks that focus on demos or data (n=13 more detailed demo, n=11 tools track, n=13 data track) might be worthwhile for upcoming venues. Evaluating the open responses to the evaluation criteria (Q18), a majority of respondents defines topics or tasks that avoid additional load on the reviewers. Some explicitly expressed this, making it obvious that good guidelines need to be created. Thus, not only reviewers but also authors should be supported when handing in research data or software.

The analysis of authors’ intentions to publish their data in the DELFI 2023 context (RQ4) raised further questions, e. g., related to demo submissions. Assuming demos present developed research prototypes, it seems as if the majority of authors at DELFI 2023 do not categorize software as research data. However, for demos presenting only the innovative usage of existing software, the authors correctly stated they have no data to publish. Furthermore, the limited availability of references to published research data in all papers at DELFI 2023 underlines the missing recognition. It is tempting to assume the lack of recognition by the community negatively impacts authors’ willingness to take on the extra work of preparing and publishing their research data. In the future, it should be evaluated how awarding open data badges for publications with published research data changes the perceived recognition and authors’ intentions to publish their research data.

To sum up, researchers’ have issues categorizing research data. They also assume and experience a variety of barriers while showing a general intent to publish their research data and software. These results underscore the need for training and guidance to support open science practices in the EdTec community.

A limitation of the presented study is the relatively low participation rate. This may partially be due to the small community, and the self-selection biases to participate or not. At the same time, it is another result of the survey, as Open Science and related practices do not seem to play a major role in the EdTec community, yet. Therefore, we consider it important to continue raising researcher's awareness, and offering support measures and incentives to engage with Open Data and Open Science practices.

6 Conclusions and Future Work

This paper presents the results of a survey study in the EdTec community that explores key areas of concern related to data ownership, sharing, and associated barriers of open science practices. We conclude that the respondents use software as well as qualitative and quantitative data by themselves and others in a broad range of formats, and recognize them as research data (RQ1). However, challenges arise when it comes to the ownership of the data. Several barriers were identified, some of a technical nature and some concerning how much data publication efforts are being recognized (RQ2). Preferred publication formats include more detailed demo and tools tracks, the enhancement of full paper tracks by the publication of the associated research data and the introduction of special data paper tracks (RQ3). Only a minority of authors published or intended to publish research data in the context of DELFI 2023 (RQ4), raising questions of missing recognition by the community and its negative impact for data publications.

Future work may entail a comparison of the results with a survey study conducted in the context of the CompEd conference [Ki23]. While the survey study was conducted in analogy, also based on [Bi23], the audience was computing education researchers worldwide [Ki24a]. The comparison may allow the identification of more barriers and issues regarding research data publications and help guide further research in support of increasing the awareness and recognition of open science practices. Furthermore, we aim to pursue the continuation of this research, e. g., by repeating the survey at the next DELFI conference, or by adding a qualitative study. For example, authors who did not publish research data along with their publications can be interviewed to elicit barriers, preferences, and intentions. Qualitative insights may also guide the design and dissemination of educational materials on open science practices for EdTec and related research communities.

Moreover, dedicated formats for the publication of research data and guiding authors during publication processes are needed as part of conferences. More networking and visibility of Open Data as a relevant topic in the EdTec community is required in the future. There is potential, for example, in the connection with other relevant actors, such as the German Research Foundation (DFG)-funded NFDIxCSS [Ge24a] project, which works towards establishing standards, providing specialized data storage offers, and community support. The latter includes guidelines for publishing data as well as supporting reviewers.

A Survey Questions

- Q1 **Categories of research data used:** Please select all the categories of research data that you used in one of your last educational technology research projects.
- Q2 **Research format(s) used:** Please select the research data formats that you used or created as part of your last educational technology research project:
- Q3 **Research data format(s) to be shared:** Which of the following research data formats you have indicated that you used or created as part of your last educational technology research project are important to be made available to others for the purpose of validation?
- Q4 **File size:** Approximately, what is the volume of the data you created and intend to share as part of your last educational technology research project? If you have none, please insert zero.
- Q5 **Number of data files:** Approximately how many files of data did you produce as part of your last educational technology research project? If you have none, please insert zero.
- Q6 **Research data ownership:** Who do you believe ‘owns’ the research data that you have made or will make available to others as part of your last educational technology research project?
- Q7 **Publishing research data:** Have you published the research data that you used or created as part of your last educational technology research project in any of the following ways?
- Q8 **Why not publish research data:** Why haven’t you published your research data?
- Q9 **Barriers to publishing data:** Thinking about your educational technology project still, has any of the following limited your ability to publish research data?
- Q10 **Barriers and concerns regarding the publication of research data:** What were the legal concerns that you had regarding publishing research data? (depends on Q9)
- Q11 **Barriers and concerns regarding the publication of research data:** What were your concerns regarding loss of control when publishing research data? (depends on Q9)
- Q12 **Barriers and concerns regarding the publication of research data:** Was one of these authority or practice considerations a reason for not publishing research data? (depends on Q9)
- Q13 **Barriers and concerns regarding the publication of research data:** Did you face one of these technical/processing barriers when publishing research data? (depends on Q9)
- Q14 **Barriers and concerns regarding the publication of research data:** Did you face one of these resource barriers when publishing research data? (depends on Q9)
- Q15 **Research project:** The educational technology research project you reported in this study ...
- Q16 **Publication formats:** What could be a possible way for publishing research data and/or research software at EdTec conferences (e. g., DELFI, LAK, EC-TEL, etc.)?
- Q17 **Mandatory upload of research data and software:** Should each of these formats be associated with the mandatory delivery of a repository, resources, etc.?
- Q18 **Review of research data and software:** What are your expectations for the review of delivered resources, repos, etc.?

References

- [Bi23] Biernacka, K.; Mulligan, A.; Zimmermann, J.; Rudiak, R.: Research Data Sharing and Reuse 2020, Online, Mendeley Data, V1, 2023, DOI: 10.17632/nr9n75cpv2.1.
- [BP21] Biernacka, K.; Pinkwart, N.: Opportunities for adopting open research data in Learning Analytics. In: *Advancing the Power of Learning Analytics and Big Data in Education*. IGI Global, Hershey, PA, pp. 29–60, 2021.
- [CB14] Clarke, V.; Braun, V.: Thematic Analysis. In: *Encyclopedia of Critical Psychology*. Springer New York, New York, NY, pp. 1947–1952, 2014, DOI: 10.1007/978-1-4614-5583-7_311.
- [Ce17] Centre for Science and Technology Studies, Elsevier and Leiden University: Open Data. The researcher perspective, tech. rep., Centre for Science and Technology Studies, Elsevier and Leiden University, 2017, URL: <https://www.elsevier.com/open-science/research-data/open-data-report>.
- [CE24] CERN European Organization for Nuclear Research: Zenodo.org, Accessed on 02/21/2024, 2024, URL: <https://www.zenodo.org>.
- [Ce24] Center for Open Science: osf.io, Accessed on 02/23/2024, 2024, URL: <https://www.osf.io>.
- [CF10] Casadevall, A.; Fang, F. C.: Reproducible Science. *Infection and Immunity* 78(12), pp. 4972–4975, 2010, DOI: 10.1128/IAI.00908-10.
- [Co24] Cornell University: arXiv.org, Accessed on 02/23/2024, 2024, URL: <https://www.arxiv.org>.
- [CRI18] Commission, E.; for Research, D.-G.; Innovation: Cost-benefit analysis for FAIR research data – Cost of not having FAIR research data. Publications Office, 2018.
- [Eu23] European Union: European Open Science Cloud, 2023, URL: <https://eosc-portal.eu/>.
- [Fa24] Fachgruppe Bildungstechnologien der Gesellschaft für Informatik e.V.: DELFI Tagungen: Themen, Formen der Forschung, Beitragsformate, Anforderungen und Einreichung, Accessed on 21.02.2024, 2024, URL: https://delfi-tagung.de/fileadmin/FG/WI-ELE/DELFI_Tagungen_Anforderungen-und-Formate.pdf.
- [FD17] Foster, E. D.; Deardorff, A.: Open science framework (OSF). *Journal of the Medical Library Association: JMLA* 105(2), p. 203, 2017.
- [Ge23a] German Centre for Higher Education Research and Science Studies (DZHW): Find Higher Education and Science Research Data Packages, 2023, URL: <https://metadata.fdz.dzhw.eu/en/start>.
- [Ge23b] German Research Foundation (DFG): Handling of Research Data, 2023, URL: https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/index.html.
- [Ge24a] German Informatics Society (GI): National Research Data Infrastructure for and with Computer Science (NFdIXCS), Accessed on 03/11/2024, 2024, URL: <https://nfdixcs.org/>.
- [Ge24b] Gesellschaft für Informatik e.V.: Call for Participation, 2024, URL: <https://delfi-tagung.de/call-for-participation>.
- [GNT10] Goecks, J.; Nekrutenko, A.; Taylor, J.: Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology* 11(8), pp. 1–13, 2010.

- [Ih15] Ihantola, P.; Vihavainen, A.; Ahadi, A.; Butler, M.; Börstler, J.; Edwards, S. H.; Isohanni, E.; Korhonen, A.; Petersen, A.; Rivers, K.; Rubio, M. Á.; Sheard, J.; Skupas, B.; Spacco, J.; Szabo, C.; Toll, D.: Educational Data Mining and Learning Analytics in Programming: Literature Review and Case Studies. In: Proceedings of the 2015 ITiCSE on Working Group Reports. ITiCSE-WGR '15, ACM, New York, pp. 41–63, 2015.
- [Ir13] Irwin, S.: Qualitative secondary data analysis: Ethics, epistemology and context. *Progress in Development Studies* 13 (4), pp. 295–306, 2013, doi: 10.1177/1464993413490479.
- [Je22] Jeuring, J.; Keuning, H.; Marwan, S.; Bouvier, D.; Izu, C.; Kiesler, N.; Lehtinen, T.; Lohr, D.; Peterson, A.; Sarsa, S.: Towards Giving Timely Formative Feedback and Hints to Novice Programmers. In: Proceedings of the 2022 Working Group Reports on Innovation and Technology in Computer Science Education. ITiCSE-WGR '22, ACM, Dublin, Ireland, pp. 95–115, 2022, doi: 10.1145/3571785.3574124.
- [KGH21] Katz, D. S.; Gruenpeter, M.; Honeyman, T.: Taking a fresh look at FAIR for research software. *Patterns* 2 (3), p. 100222, 2021, doi: 10.1016/j.patter.2021.100222.
- [Ki22a] Kiesler, N.: Dataset: Recursive problem solving in the online learning environment CodingBat by computer science students, Online, Datenerhebung: 2017. Version: 1.0.0. Datenpaketzugangsweg: Download-SUF. Hannover: FDZ-DZHW, 2022, doi: 10.21249/DZHW:studentsteps:1.0.0.
- [Ki22b] Kiesler, N.: Daten- und Methodenbericht Rekursive Problemlösung in der Online Lernumgebung CodingBat durch Informatik-Studierende, tech. rep., DZHW, 2022, url: [https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-studentsteps\\$/attachments/studentsteps_Data_Methods_Report_de.pdf](https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-studentsteps$/attachments/studentsteps_Data_Methods_Report_de.pdf).
- [Ki23] Kiesler, N.; Impagliazzo, J.; Biernacka, K.; Kapoor, A.; Kazmi, Z.; Ramagoni, S. G.; Sane, A.; Tran, K.; Taneja, S.; Wu, Z.: Where's the Data? Exploring Datasets in Computing Education. In: Proceedings of the ACM Conference on Global Computing Education Vol 2. CompEd 2023, ACM, Hyderabad, India, pp. 209–210, 2023, doi: 10.1145/3617650.3624951.
- [Ki24a] Kiesler, N.; Impagliazzo, J.; Biernacka, K.; Kapoor, A.; Kazmi, Z.; Ramagoni, S. G.; Sane, A.; Tran, K.; Taneja, S.; Wu, Z.: Where's the Data? Finding and Reusing Datasets in Computing Education. In: Proceedings of the Working Group Reports on Global Computing Education. CompEd 2023, ACM, Hyderabad, India, 2024.
- [Ki24b] Kiesler, N.; Röpke, R.; Schiffner, D.; Schulz, S.; Strickroth, S.; Ehlenz, M.; Heinemann, B.; Wilhelm-Weidner, A.: Towards Open Science at the DELFI Conference, 2024, doi: 10.17605/OSF.IO/VE5DB, url: <https://osf.io/ve5db>.
- [KPC21] Katz, D. S.; Psohopoulos, F. E.; Castro, L. J.: Working Towards Understanding the Role of FAIR for Machine Learning. In: 2nd Workshop on Data and Research Objects Management for Linked Open Science. ZB MED-Publikationsportal Lebenswissenschaften, Online, pp. 1–6, 2021, doi: 10.4126/FRL01-006429415.
- [KS22] Kiesler, N.; Schiffner, D.: On the Lack of Recognition of Software Artifacts and IT Infrastructure in Educational Technology Research. In (Henning, P. A.; Striewe, M.; Wölfel, M., eds.): 20. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 201–206, 2022, doi: 10.18420/de1fi2022-034.
- [KS23a] Kiesler, N.; Schiffner, D.: Open Science in den Bildungstechnologien: Zur Publikation und Begutachtung von Forschungsdaten inklusive Software im Rahmen der DELFI. In: Workshops der 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 159–168, 2023, doi: 10.18420/wsdel1fi2023-40.

- [KS23b] Kiesler, N.; Schiffner, D.: Why We Need Open Data in Computer Science Education Research. In: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education Vol. 1. ITiCSE 2023, ACM, Turku, Finland, pp. 348–353, 2023, DOI: 10.1145/3587102.3588860.
- [Mc19] McGill, M. M.: Discovering Empirically-Based Best Practices in Computing Education Through Replication, Reproducibility, and Meta-Analysis Studies. In: Proceedings of the 19th Koli Calling International Conference on Computing Education Research. Koli Calling '19, ACM, Koli, Finland, 2019, DOI: 10.1145/3364510.3364528.
- [Mo11] Mons, B.; van Haagen, H.; Chichester, C.; Hoen, P.-B.; den Dunnen, J. T.; van Ommen, G.; van Mulligen, E.; Singh, B.; Hooft, R.; Roos, M., et al.: The value of data. *Nature genetics* 43 (4), pp. 281–283, 2011.
- [MS23] Melkamu Jate, W.; Striewe, M.: Umgang der DELFI-Community mit Forschungsdaten und Softwareartefakten - Eine Erhebung auf Basis der Tagungsbände im Zeitraum 2018-2022. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 167–172, 2023, DOI: 10.18420/delfi2023-27.
- [Na13] National Science Foundation: Open Data at NSF, 2013, URL: <https://www.nsf.gov/data/>.
- [Pa02] Patton, M. Q.: *Qualitative Research & Evaluation Methods*. Sage, Thousand Oaks, 2002.
- [PF07] Pilat, D.; Fukasaku, Y.: OECD principles and guidelines for access to research data from public funding. *Data Science Journal* 6, OD4–OD11, 2007.
- [Pr23] Prather, J.; Denny, P.; Leinonen, J.; Becker, B. A.; Albluwi, I.; Craig, M.; Keuning, H.; Kiesler, N.; Kohn, T.; Luxton-Reilly, A.; MacNeil, S.; Petersen, A.; Pettit, R.; Reeves, B. N.; Savelka, J.: The Robots Are Here: Navigating the Generative AI Revolution in Computing Education. In: Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education. ITiCSE-WGR '23, ACM, Turku, Finland, pp. 108–159, 2023, DOI: 10.1145/3623762.3633499.
- [SBH23] Schulz, S.; Berndt, S.; Hawlitschek, A.: Gruppenarbeit beim Programmieren lernen (GAPL). Datenerhebung: 2020. Version: 1.0.0. Datenpaketzugangsweg: SUF: Download, ed. by İkiz-Akinci, D.; Niebuhr, J., Hannover, 2023, DOI: 10.21249/DZHW:dipit2020:1.0.0.
- [SS22] Strickroth, S.; Striewe, M.: Building a Corpus of Task-based Grading and Feedback Systems for Learning and Teaching Programming. *International Journal of Engineering Pedagogy (IJEP)* 12 (5), pp. 26–41, 2022, DOI: 10.3991/ijep.v12i5.31283.
- [Wi16] Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E., et al.: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3 (1), pp. 1–9, 2016.