

# Dublin Core for Recording Metadata of Experiments in Software Engineering: A Survey

Filipe A. Santana, André F. R. Cordeiro, Edson Oliveira Jr

<sup>1</sup>Departamento de Informática – Universidade Estadual de Maringá (UEM)  
Maringá – PR – Brazil

ra123515@uem.br, cordeiroandrefelipe@gmail.com, edson@din.uem.br

***Abstract.** Metadata is defined as data about data. Metadata Standards represent structures for recording in order to promote data interoperability. This study focuses on investigating the Utility and Feasibility of the Dublin Core Standard for recording metadata associated with Software Engineering Experiments. A Survey was conducted with professionals and researchers who have experience. Eleven answers were collected. The observed results suggest positive research prospects.*

## 1. Introduction

Metadata can be defined as data that describes, explains, or references other data [NCES 2023]. Metadata standards are structures that represent the characteristics of a set of metadata [Cardoso 2020]. Among the benefits associated with metadata standards are data retrieval and interoperability [Hayslett 2023]. The literature presents various standards for metadata representation. Notable examples include Dublin Core (DC) [DCMI 2023], ADL Scorm [SCORM 2023], Motion Imagery Standards Board (MISB) [MISB 2023], IEEE LOMv1.0 [IEEE 2023], and Resource Description Framework (RDF) [W3C 2023].

The diversity of standards can be analyzed based on different criteria such as documentation, adaptability, automated processing, and popularity. In this context, DC stands out for its extensive documentation [DCMI 2020], adaptability for different domains [Clayphan and Guenther 2004], feasibility of automated processing [DCMI 2019], and popularity [Higgins 2007].

The features mentioned for DC influenced the choice of this standard for recording metadata collected and/or produced in activities carried out in Software Engineering (SE). A survey [Guidolin et al. 2021] was conducted with professionals and researchers experienced in Experimental Software Engineering (ESE) to assess the Utility and Feasibility of DC for metadata recording in SE.

The following sections of this study provide more information. In Section 2, the Research Methodology is described. In Section 3, the research objectives and questions are presented. In Sections 4 and 5, information about the target audience and the sample considered in the survey is provided, respectively. In Sections 6 and 7, details about the assessment instrument and data sharing are presented. In Sections 8 and 9, the results are presented and discussed.

## 2. Research Methodology

This study is exploratory in nature and aims to deepen our understanding of the perception of professionals and researchers with experience in ESE regarding the Utility and Feasibility of DC for recording experiments in SE. A survey was conducted in this study. Surveys are conducted to collect data from a sample of individuals [Scheuren 2023], in the most reliable and unbiased manner possible [Guidolin et al. 2021].

The DC was chosen due to its use in various fields, such as librarianship [Clayphan and Guenther 2004], availability of documentation [DCMI 2020], automated processing [DCMI 2019], and flexibility [Higgins 2007]. The standard allows for the recording of metadata in terms of Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights [Higgins 2007].

Initially, three conducted experiments were described using DC and were considered during data collection. Data collection was carried out through a questionnaire sent to professionals and researchers. Among the groups of questions considered were those associated with Professional Profile, Utility, and Viability of DC usage.

## 3. Goals and Research Questions

To enhance the understanding of the Utility and Feasibility of DC for recording experimental metadata in SE, a questionnaire was administered to professionals and researchers with experience in ESE. The questionnaire included various questions, and the following subsections present the question groups.

### 3.1. Participant Characterization

The characterization questions are presented with the aim of understanding the participants' level of experience. In this study, the following characterization questions were considered:

- **QCP1:** What is your educational level?
- **QCP2:** What is your area of specialization in Software Engineering?
- **QCP3:** Which best describes your professional activities?
- **QCP4:** How long have you been working in Software Engineering Experimentation?
- **QCP5:** What is your experience with controlled experimentation in Software Engineering?
- **QCP6:** Are you familiar with any metadata standards for any purpose?
- **QCP7:** Are you familiar with the Dublin Core metadata standard?
- **QCP8:** Have you used the Dublin Core metadata standard for any reason?

The first question aims to understand the educational background of the participants. The second question seeks to understand the specialization or focus of the participants within the field of SE which is relevant to analyze the use of DC. The third question is designed to provide a comprehensive view of the tasks performed by the participants in their careers, highlighting their specific responsibilities in SE.

In the fourth question, the interest is in the assessment of the participants' experience in ESE and their level of familiarity with the subject and it is also relevant to

analyze the use of DC. The fifth question, **"What is your experience with controlled experimentation in Software Engineering?"** focuses on the participants' experience with controlled experiments, a fundamental element in experimental research. The sixth question investigates the participants' general knowledge of metadata standards, regardless of the context of SE. The seventh question is specifically directed to assess the participant's knowledge of the DC standard.

Finally, in the eighth question, **"Have you used the Dublin Core metadata standard for any reason?"** explores the practical experience of the participants with the use of DC.

### 3.2. Questions on Dublin Core

The scientific questions are presented with the aim of understanding the perception of professionals and researchers regarding the use of DC for recording experimental metadata in SE in terms of Utility and Feasibility. In this study, the following scientific questions were considered:

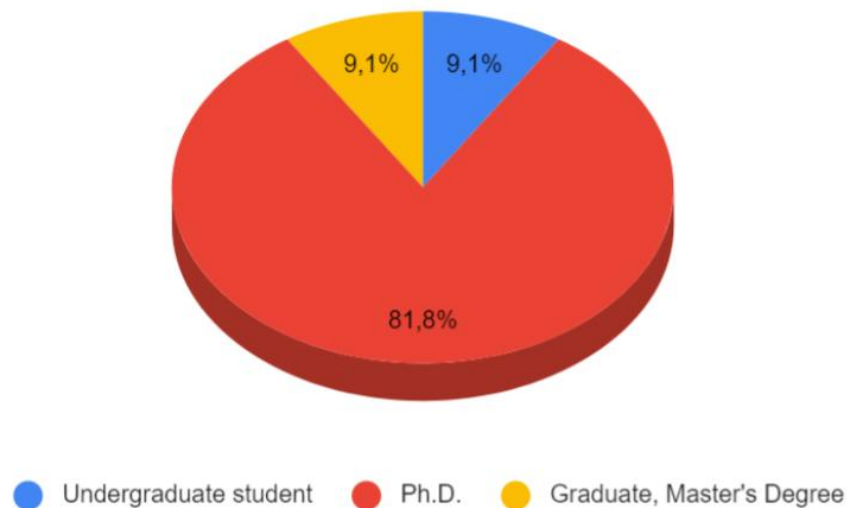
- **QC1:** Can Dublin Core be useful for describing metadata of controlled experiments in Software Engineering?
- **QC2:** Is it feasible to use Dublin Core to describe metadata of controlled experiments in Software Engineering?
- **QC3:** Would you consider using Dublin Core to record metadata related to controlled experiments in Software Engineering?
- **QC4:** Are you aware of any competing metadata standards for controlled experiments?
- **QC5:** Considering what has been presented, would you like to suggest possible improvements?

Dublin Core consists of 15 basic elements that assist in the retrieval of digital resources. It is one of the simplest and most widely used metadata standards, offering comprehensive data description and organization, automated processing, and complete documentation [DCMI 2023].

The question QC1 aims to assess participants' perception of the utility of the DC standard in describing metadata for controlled experiments. This will help determine whether professionals consider DC as a relevant tool in this context. The second question, **"Is it feasible to use Dublin Core to describe metadata for controlled experiments in Software Engineering?"**, seeks to understand whether participants believe that the application of DC is viable for describing metadata of controlled experiments in SE. The third question explores participants' willingness to adopt DC for recording metadata related to controlled experiments, considering their preference for this standard.

The fourth question aims to investigate participants' knowledge of other metadata standards for controlled experiments. The fifth question, **"Considering what has been presented, would you like to suggest possible improvements?"** offers participants the option to contribute with feedback to enhance the use of DC in the context of experiment metadata in SE.

What is your educational level?



**Figure 1. Participants' Educational level**

#### **4. Target Audience, Population and Sampling**

The need to assess DC in terms of Utility and Feasibility required a specific participant profile. For this study, it was considered that each participant should have experience in ESE, specifically in conducting experiments and replications. More information about the participants' profiles is presented in the results section.

#### **5. Instrument and Evaluation**

The questionnaire used in this study was initially evaluated by researchers who fit the participant profile. The improvements identified in the evaluation were addressed. Subsequently, the questionnaire was made available for 30 days for professionals and researchers to participate. The questions presented in subsections 3.1 and 3.2 make up the questionnaire.

#### **6. Data Sharing**

Data of this study is available at <https://zenodo.org/records/10041670>.

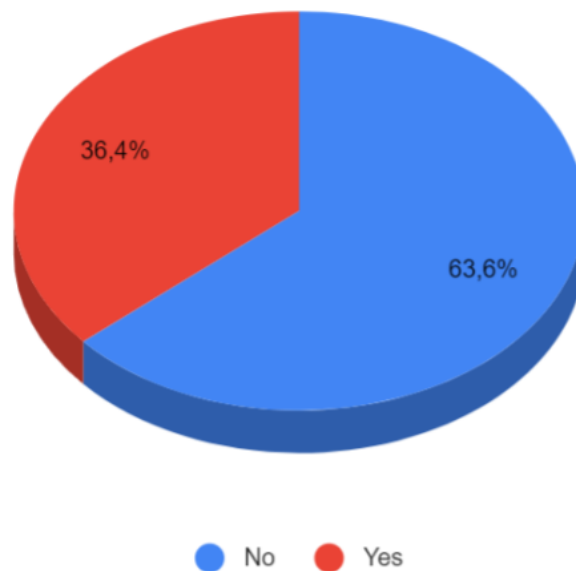
#### **7. Results**

The results obtained in the study are presented in this section. These results have been grouped in terms of participant characterization and questions about DC.

##### **7.1. Participant Characterization**

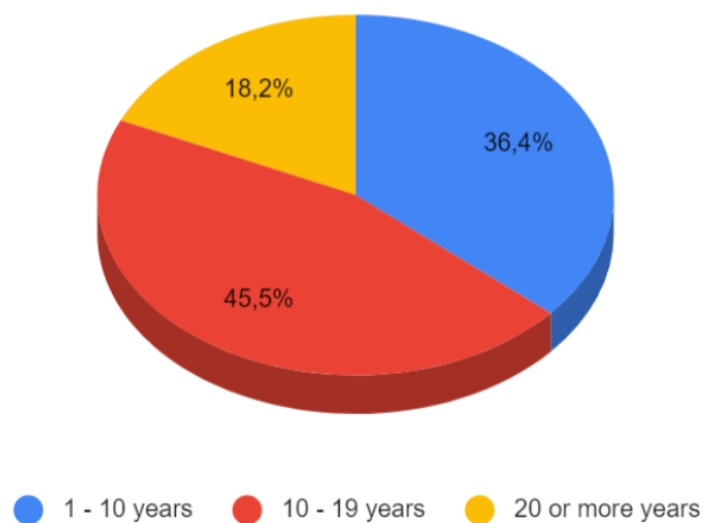
The results regarding the participant characterization profile can be observed in Figures 1, 2, and 3.

Do you know about the Dublin Core or any other metadata standard for any purpose?



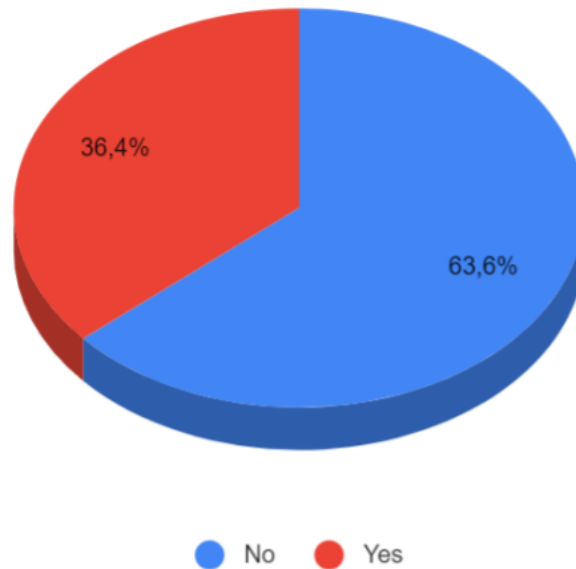
**Figure 2. Years of Experience in ESE**

How long have you been working with Software Engineering Experimentation?



**Figure 3. Experience level with experimentation in SE**

Do you know about the Dublin Core or any other metadata standard for any purpose?



**Figure 4. Knowledge about metadata standards and DC**

Considering Figures 1, 2, and 3, it was observed that out of the 11 collected responses, over 80% of the researchers hold a Ph.D., 9% have a master's degree, and 9% are undergraduates. It is also observed that more than 60% of the participants have over 10 years of experience in the field of SE, and over half of the researchers possess knowledge of the experimental process, execution, and replication of experiments in SE.

Regarding knowledge about metadata, it is noted that approximately 40% of the participants are familiar with Metadata Standards. Among these participants, all stated that they were acquainted with DC. More than 70% of the researchers believe that DC can be useful and feasible for describing experiment metadata in SE and would consider using the standard. Approximately 20% of the researchers are familiar with some competing metadata standards for controlled experiments. Notable examples include the Asset Description Metadata Schema (ADMS) metadata standard.

## **7.2. Questions on Dublin Core**

The results related to the questions presented in section 3.2 can be observed in Table 1.

Considering the results presented in Table 1, it can be observed that for QC1, **"Have you ever used the Dublin Core metadata standard for anything?"** 36% of the participants stated that they had not used DC, 28% stated that they had used DC, and 36% did not respond. For QC2, **"Can DC be useful for describing metadata of experiments in Software Engineering?"** 64% agreed with the statement, 9% fully agreed, 9% remained neutral, and 18% disagreed with the statement. In the case of QC3, **"Is it feasible to use DC to describe metadata of experiments in Software Engineering?"**

**Table 1. Questions on Dublin Core**

<b>Participants</b>	<b>QC1</b>	<b>QC2</b>	<b>QC3</b>	<b>QC4</b>	<b>QC5</b>
Participant 1	Neutral	Strongly Agree	Strongly Agree	Yes	No
Participant 2	Neutral	Disagree	Strongly Agree	Yes	No
Participant 3	Yes	Agree	Agree	Yes	No
Participant 4	No	Neutral	Neutral	No	Yes
Participant 5	No	Disagree	Disagree	No	No
Participant 6	Yes	Agree	Agree	Yes	No
Participant 7	Yes	Agree	Strongly Agree	Yes	No
Participant 8	No	Agree	Agree	No	Yes
Participant 9	Neutral	Agree	Agree	Yes	No
Participant 10	No	Agree	Agree	Yes	No
Participant 11	Neutral	Agree	Strongly Agree	Yes	No

27% responded that they fully agree, 45% responded that they agree, 9% remained neutral, and 9% disagreed. Regarding QC4, **"Would you consider using the Dublin Core to record metadata related to controlled experiments in Software Engineering?"** 73% responded "Yes," and 27% responded "No." As for QC5, **"Are you familiar with any competing metadata standards for controlled experiments?"** 82% of the participants answered "No," and 18% answered that they are familiar with some competing metadata standards for controlled experiments.

Among the mentioned standards is the Asset Description Metadata Schema (ADMS) metadata standard.

## **8. Discussion of Results**

Based on the participant characterization questionnaire, a variety of experiences in the field of SE can be observed. The authors understand that such diversity contributes to the understanding of the scenario related to metadata usage. In terms of the results obtained regarding the questions from the DC questionnaire.

For the QC1, one-third of the participants lack familiarity or experience with the DC, while 28% have already used the DC, indicating they have a certain degree of metadata standard experience, and one-third of the participants chose not to answer. In the QC2, the majority of participants agreed that DC can be useful for describing metadata of experiments in SE. 18% of the participants disagreed with the statement, which suggests there may be issues or preferences for other possible alternatives.

In the case of the QC3, the majority of participants see DC as a viable option, with 45% of participants partially agreeing and 27% fully agreeing. The remaining 9% who stayed neutral and 9% who disagreed suggest that some participants believe that DC may not be a viable option. For the QC4, the vast majority of participants, approximately a third, expressed their willingness to adopt DC, while 27% do not intend to use it. This may represent those who have preferences for other standards or who are not familiar with DC.

In QC5, most participants are not familiar with other competing standards for describing metadata in SE experiments, potentially indicating that participants do not see

the need to explore alternatives to DC. Meanwhile, 18% of the participants are familiar with a competing standard, ADMS, being one of them.

Based on the results obtained, it is possible to observe that, despite the study's specific subject, the majority of responses (80%) were positive regarding the use of DC to describe metadata of experiments in SE.

## 9. Prospective Actions and Final Remarks

This study presents a survey on the use of the DC Metadata Standard for recording experimental metadata in SE. A sample composed of 11 assessments was constructed. From this sample, it was possible to observe results that indicated potential. Such evidence is related to Utility and Feasibility.

The authors acknowledge that this evidence is preliminary due to the number of participants. Therefore, the replication of this survey is planned for future studies to expand the sample size. Another future study to be conducted is related to the development of a tool for recording metadata specifications using DC in the context of SE experiments.

## Acknowledgements

Edson Oliveira Jr thanks CNPq/Brazil Grant #311503/2022-5.

## References

- Cardoso, I. P. (2020). Dados científicos e metadados: estudo sobre o uso dos padrões de metadados no fluxo da informação científica sobre biodiversidade.
- Clayphan, R. and Guenther, R. (2004). Library application profile - draft. <https://www.dublincore.org/specifications/dublin-core/library-application-profile/>.
- DCMI (2019). Dcmi community groups. <https://www.dublincore.org/groups/>.
- DCMI (2020). Dcmi metadata terms. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>.
- DCMI (2023). Dublin core. <https://www.dublincore.org/specifications/dublin-core/>.
- Guidolin, K., Wexner, S. D., Jung, F., Khan, S., Deng, S. X., Kirubarajan, A., Qureshy, F., and Chadi, S. (2021). Strengths and weaknesses in the methodology of survey-based research in surgery: A call for standardization. *Surgery*, 170(2):493–498.
- Hayslett, M. (2023). Metadata for data management: A tutorial: Intro. <https://guides.lib.unc.edu/metadata>.
- Higgins, S. (2007). What are metadata standards. <https://www.dcc.ac.uk/guidance/briefing-papers/standards-watch-papers/what-are-metadata-standards>. Accessed on: 29 Jul. 2023.
- IEEE, S. f. L. O. M. (2023). Learning technology standards committee of the IEEE computer society. <https://ieeexplore.ieee.org/document/9262118>.
- MISB, G.-I. S. W. G. (2023). Motion imagery standards board (misb). [https://gwg.nga.mil/gwg/focus-groups/Motion\\_Imagery\\_Standards\\_Board\\_\(MISB\).html](https://gwg.nga.mil/gwg/focus-groups/Motion_Imagery_Standards_Board_(MISB).html).



NCES (2023). National center for education statistics - nces. forum guide to metadata the meaning behind education data. [https://nces.ed.gov/pubs2009/metadata/ch1\\_overview.asp#r1](https://nces.ed.gov/pubs2009/metadata/ch1_overview.asp#r1).

Scheuren, F. (2023). What is a survey? <https://www.unh.edu/institutional-research/sites/default/files/media/2022-05/what-is-a-survey.pdf>.

SCORM (2023). Sharable content object reference model. <https://adlnet.gov/past-projects/scorm/#scorm-versions-and-resources>.

W3C (2023). Resource description framework (rdf) model and syntax specification. <https://www.w3.org/TR/PR-rdf-syntax/>.