Use of the Dublin Core Standard to Express Open Metadata Related to Software Engineering Experiments

Filipe Amadeu Santana
Informatics Department
State University of Maringá
Maringá, Brazil
ra123515@uem.br

André F. R. Cordeiro
Informatics Department
State University of Maringá
Maringá, Brazil
cordeiroandrefelipe@gmail.com

Edson OliveiraJr
Informatics Department
State University of Maringá
Maringá, Brazil
edson@din.uem.br

ABSTRACT

Metadata represents descriptive information about data. They facilitate data retrieval and ensure data interoperability. Metadata can be expressed with metadata standards. In this study, we investigated using metadata standards to express metadata of Software Engineering (SE) Experiments. We compared five different standards. The Dublin Core Standard presented better results based on the documentation criteria, automated processing, and the possibility of application in an SE experimentation context. We provide examples of how to use Dublin Core. In a future study, a tool will be developed to perform metadata specifications in Dublin Core toward promoting open science practices for SE experiments.

KEYWORDS

Metadata, Metadata Standard, Dublin Core, Software Engineering Experiments.

INTRODUCTION

Metadata refers to data about data (RILEY, 2017). It provides information about the data, making data retrieval easier (RILEY, 2017). Metadata standards act as organizing frameworks, optimizing metadata retrieval and interoperability (HAYSLETT, 2023). Among the metadata standards, Dublin Core is one of the simplest and most widely used, offering comprehensive data description and organization, enabling automated processing, and possessing complete documentation (DCMI, 2023).

This study aims to explore metadata standards that can be adapted to describe metadata in Software Engineering (SE) experiments. Considering the possible benefits associated with the use of the metadata, we understand that metadata can facilitate the reproduction of experiments, in terms of data and metadata. The study considered as initial stage of a literature review and an evaluation of the metadata standards. At the end of the evaluation, we chose the Dublin Core Standard Metadata to describe metadata related to experiments carried out in SE. The Dublin Core metadata standard is one of the most widely used standards (DCC, 2023). This characteristic suggests a flexibility that could be tested to register metadata related to SE experiments.

BACKGROUND

The term metadata refers to data about data, descriptive information that provides context to a given piece of data (HAYSLETT, 2023). According to GRÁCIO (2012), metadata is a set of essential elements for managing data and enabling easy retrieval of this data. The number of metadata elements varies depending on the adopted standard. Metadata can be applied in various scientific fields, not limited to computer science (FORMENTON et al., 2018). However, for metadata to become usable, they must be standardized, ensuring data interoperability (HAYSLETT, 2023).

Metadata standards can be seen as guidelines or structures used to organize metadata in a way that supports data retrieval and communication between these data (NISO, 2017). They can be divided into: **Descriptive Metadata Standards**, which contain elements like title, author, subject, description, and others; **Structural Metadata Standards**, which establish hierarchies between digital elements (text, image, video, audio); **Administrative Metadata Standards**, which help manage archival resources; **Technical Metadata Standards, and Preservation Metadata Standards** (FORMENTON et al., 2018). There are various other metadata standards with specific purposes. Each standard plays an important role in data organization and retrieval (NISO, 2017).

There is an extensive list of metadata standards for various formats and types of data, such as Access to Biological Collections Data (ABCD), and Agricultural Metadata Element Set (AgMES) (DCC, 2023). The ABCD is used to access and exchange data that document where and when species were registered. The AgMES is a semantic metadata standard used to describe numerous types of informational resources relevant to the food production industry (DCC, 2023). There is also the Dublin

Core, a simple metadata standard consisting of basic elements easily implemented, besides being one of the most widely used standards (DCC, 2023).

METHODOLOGY

Initially, it was necessary to conduct a literature review, performed by the author and reviewed by the co-author, Master André F. R. Cordeiro, alongside Professor Edson Oliveira Jr. A protocol was defined. The following information was considered: research objective, research questions, search strategy, bibliographic bases, search string, and inclusion and exclusion criteria for choosing the standard. The details related to the protocol are presented in Table 1. Considering the information presented in Table 1, it is evident our interest in investigating metadata standards used or applied in Computer Science. Through the exploration of these standards, we would select one to be considered within the context of experimentation in Software Engineering.

Table 1: Description of the Literature Review Protocol.

Literature Review Protocol	
Topic	**Description**
Research Objective	Identify metadata standards used in one or more areas of Computer Science.
Research Questions	What metadata standards have been used in one or more areas of Computer Science? What areas of Computer Science have considered the use of metadata standards?
Search Strategy	Automatic search in computer science bibliographic databases.
Bibliographic Bases	IEEE Xplore; ACM Digital Library; SpringerLink; ScienceDirect
Search String	"computer science" AND (“metadata” or “metadata standards”)
Inclusion Criteria	I1 - Explicit citation of one or more metadata standards that are being used in one or more areas of Computer Science. I2 - Explicit citation of one or more areas of Computer Science.
Exclusion Criteria	E1 - Study that is not inserted in the context of Computer Science. E2 - Non-explicit citation of one or more metadata standards that are being used in one or more areas of Computer Science. E3 - No explicit citation of one more areas of Computer Science. E4 - Study not written in English (difficults dissemination and reproducibility). E5 - Studies that have not been published in conferences or journals. E6 - Duplicate studies. E7 - Studies unavailable, even after contacting the authors. E8 - Secondary studies.

After the literature review, the analysis of the found metadata standards was started. A summary of the previously gathered standards was prepared, including the title of the study, a description of the study, the field of computer science considered in the research, and the description of the metadata standard usage. After completing these steps, the selection of five candidate metadata standards for the context of the experiments in software engineering was carried out, along with the justification for their choice. The five candidate metadata standard we selected are Dublin Core (https://www.dublincore.org); ADL SCORM (Sharable Content Object Reference Model (SCORM®) | ADL Initiative (adlnet.gov)); Motion Imagery Standards Board (NSG Documents Registry Search Results (nga.mil)); IEEE LOMv1.0 (1484.12.1-2020 - IEEE Standard for Learning Object Metadata | IEEE Standard | IEEE Xplore); RDF (Resource Description Framework (RDF) Model and Syntax Specification (w3.org)).

In this study, we decided to use the Dublin Core Metadata Standard to document metadata related to Software Engineering Experiments. The choice of Dublin Core was based on the comparison between metadata standards mentioned previously. The criteria used in the comparison were the presence of complete documentation, automated processing, and a possible flexibility for application in the context of Software Engineering. Table 2 presents the results related to the comparison.

Table 2: Comparison between the Metadata Standards.

Metadata Standard	Complete Documentation	Automated Processing	Flexibility for Software Engineering
Dublin Core	X	X	X
ADL SCORM	X	X	X
Motion Imagery Standard Board	X	X	
IEEE LOMv1.0	X	X	X
RDF	X	X	X

Although it may seem like ADL SCORM, IEEE LOM, and RDF could be chosen instead, the Dublin Core offers more usage examples, even from third parties, and it would not require much adaptation, unlike the other metadata standards.

Before applying the Dublin Core in examples of experiments, the experimental process in Software Engineering, described in WOHLIN et al. (2012), was studied. It was necessary to map the elements of Dublin Core (DCMI, 2019) to each of the elements of the experimental process in Software Engineering to make the specification in Dublin Core viable for the selected experiments.

EXAMPLES OF USE

To express the use of Dublin Core, we selected experiments reported in the software engineering literature. We selected a few experiments from the book about Software Engineering Experimentation (WOHLIN et al., 2000). These experiments were chosen because they are associated with the application of an experimental process, described in the book. We aimed to keep the specification of the experiments in Dublin Core as concise as possible, however, there were still repetitions in some cases. The main features of the experiments are presented in the examples below.

In the first experiment (The Maryland-95 Study) (PORTER; VOTTA; BASILI, 1995), an investigation was conducted at the University of Maryland to test the effectiveness of the Defect-Based Reading (DBR) method in detecting different classes of problems. Table 3 expresses possible metadata related to this experiment.

Table 3: DBR-01 – “Comparing Detection Methods for Software Requirements Inspections: A Replicated Experiment”

Experiment Definition	
Metadata Element	**Use**
Title	“Comparing Detection Methods for Software Requirements Inspections: A Replicated Experiment”
Subject	“The methods used to perform fault detection”
Description	“Conduction of a multi-trial experiment to characterize the behavior of existing approaches and evaluate the potential benefits of scenario-based methods. The hypothesis of the experiment is that a scenario-based method, in which each reviewer uses different systematic techniques to look for different specific classes of failures, will have a significantly higher success rate”
Creator	“Adam A. Porter, Lawrence G. Votta Jr, Victor R. Basili”
Publisher	“IEEE”
Contributor	“Mark Ardis, John Kelly, David Weiss, John Gannon, Richard Gerber, Clive Loader, Eric Slud, Scott VanderWeil, Art Caso and the 48 Computer Science students”
Date	“June 1995”
Type	“Text”
Format	“Portable Document Format (PDF)”
Identifier	“DBR-01”
Source	“IEEE database”
Language	“English”
Relation	“Fault detection method”
Coverage	"Available in IEEE Digital Library"
Rights	“The rights are associated with the database's copyrights”

The second experiment deals with the Perspective-Based Reading (PBR) method for error detection in an industrial setting. The experiment was conducted with software developers from the National Aeronautics and Space Administration (NASA), at the Goddard Space Flight Center Laboratory (BASILI et al., 1996). Table 4 expresses the possible metadata related to this experiment.

Table 4: PBR-01 - “The Empirical Investigation of Perspective-Based Reading”

Experiment Definition	
Metadata Element	Use
Title	“The Empirical Investigation of Perspective-Based Reading”
Subject	“Perspective-Based Reading (PBR)”
Description	“This paper deals with reading techniques known as Perspective- Based Reading (PBR), and its application to requirements documents. The goal of PBR is to provide operational scenarios where members of a review team read a document from a particular perspective”
Creator	“Victor R. Basili, Scott Green, Oliver Laitenberger, Filippo Lanubile, Forrest Shull, Sivert Sorumgard, Marvin V. Zelkowitz”
Publisher	“Springer”
Contributor	“26 professional software developers from the National Aeronautics and Space Administration /Goddard Space Flight Center”
Date	“1996”
Type	“Text”
Format	“Portable Document Format (PDF)”
Identifier	“PBR-01”
Source	“SpringerLink Database”
Language	“English”
Relation	“Perspective-Based Reading Technique”
Coverage	"Available in SpringerLink Database"
Rights	“1996 Kluwer Academic Publishers, Boston”

The third experiment also focuses on the DBR method, but this time, it involves replication of the first experiment. It was carried out at Linkoping University in Sweden (SANDAHL et al, 1998). The experiment involved the collaboration of students, however, their specific courses were not specified. Table 5 expresses possible metadata related to this experiment.

Table 5: BDR-04 - “An Extended Replication of an Experiment for Assessing Methods for Software Requirements Inspections”

Experiment Definition	
Metadata Element	**Use**
Title	“An Extended Replication of an Experiment for Assessing Methods for Software Requirements Inspections”
Subject	“Comparing the Scenario method and the Checklist method for inspecting requirements specifications.”
Description	“The experiment has been conducted in an educational context. The study manipulated three independent variables: detection method, requirements specification, and the order of the inspections. The dependent variable measured is the defect detection rate. The experiment is relevant in the context of validate different types of defect detection techniques such as Ad Hoc, Checklist and Defect-based Scenario in an educational context”
Creator	“Kristian Sandahl, Ola Blomkvist, Joachim Karlsson, Christian Krysander, Mikael Lindvall, Niclas Ohlsson”
Publisher	“Springer”
Contributor	“24 Computer Science students”
Date	“1998”
Type	“Text”
Format	“Portable Document Format (PDF)”
Identifier	“DBR-04”
Source	“SpringerLink”
Language	“English”
Relation	“Porter, et al.: Comparing detection methods for software requirements inspection: a replicated experiment”
Coverage	"Available in SpringerLink Database"
Rights	“1996 Kluwer Academic Publishers, Boston”

In Tables 3, 4, and 5 of each experiment, all 15 elements associated with the Dublin Core Standard were considered. From these elements, it was possible to register the metadata. The authors understand that from the metadata collected and presented, it is possible to understand the context of each study associated with each experiment.

Tables 3, 4, and 5 present metadata from an experiment overview. The authors observe a second possibility, related to the registration of metadata for the activities of the experimental process, such as planning, operation, and analysis (WOHLIN et al., 2012). Despite the initial observation, further investigations need to be carried out to assess the viability of this type of record.

Considering the possible benefits associated with the use of metadata, we understand that metadata can contribute to the reproduction of experiments in software engineering, mainly when combined with experimental packages (SOLARI e VEGAS, 2006).

For more information about Dublin Core, see its documentation (Dublin Core TM, 2023).

LIMITATIONS OF THE STUDY

This study focused only on the investigation of a metadata standard to be used in software engineering experiments. In this context, we came to the conclusion of choosing the Dublin Core metadata standard. The implementation of this standard in a tool will be considered in future studies.

FUTURE STUDIES

We have some plans for future studies related to this research. We will conduct one survey to gather evaluations from experienced researchers in the field of Software Engineering Experimentation, about the use of Dublin Core for metadata records, in terms of usefulness and feasibility.

Another plan is related to the development of a tool to enable metadata specifications with Dublin Core, in the context of Software Engineering Experiments.

CONCLUSION

Throughout this study, we show the possibility of using one metadata standard to express metadata generated about software engineering experiments. We believe that Dublin Core can be applied to simplify the documentation of experimental data, in terms of the registration of metadata. Three examples of use were presented to demonstrate this possibility.

In this study, we focused on an investigation of a metadata standard that could be used to record metadata related to software engineering experiments. In future studies, we will conduct a survey to validate the usefulness and feasibility of Dublin Core. Additionally, we will develop a tool to perform Dublin Core specifications.

In conclusion, the Dublin Core Metadata Standard proved to be a possibility for describing and organizing metadata in Experimental Software Engineering. Its application has shown to be a viable option, enhancing the management, retrieval, and interoperability of experimental data.

REFERENCES

BASILI, V. R. et al. The empirical investigation of Perspective-Based Reading. Empirical Software Engineering, v. 1, n. 2, p. 133–164, 1996. Accessed on: 29 jul. 2023.

DCMI: About DCMI. Available at: <https://www.dublincore.org/about/>. Accessed on: 29 jul. 2023.

DCMI: DCMI Metadata Terms. Available at: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>. Accessed on: 29 jul. 2023.

Dublin Core TM. Available at: <https://www.dublincore.org/specifications/dublin-core/>. Accessed on: 29 jul. 2023.

FORMENTON, D. et al. Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital. Biblios: Journal of Librarianship and Information Science, n. 68, p. 82–95, 12 jan. 2018.

GRÁCIO, J. C. A. UNESP. Preservação digital na gestão da informação: um modelo processual para as instituições de ensino superior. http://hdl.handle.net/11449/144956, p. 214 p., 2012.

GWG MOTION IMAGERY STANDARDS BOARD (MISB). FG - Motion Imagery Standards Board. Available at: <https://nsgreg.nga.mil>. Accessed on: 26 jul. 2023.

HAYSLETT, M. LibGuides: Metadata for Data Management: A Tutorial: Intro. Disponível em: <https://guides.lib.unc.edu/metadata>. Accessed on: 29 jul. 2023.

HIGGINS, Sarah. What are Metadata Standards Aberystwyth University, February 2007. Available at: https://www.dcc.ac.uk/guidance/briefing-papers/standards-watch-papers/what-are-metadata-standards. Accessed on: 29 jul. 2023.

List of Metadata Standards | DCC. Available at: <https://www.dcc.ac.uk/guidance/standards/metadata/list?page=1>. Accessed on: 29 jul. 2023.

Learning Technology Standards Committee Of The IEEE Computer Society. IEEE Standard for Learning Object Metadata. Disponível em: <https://ieeexplore.ieee.org/document/9262118>. Accessed on: 29 jul. 2023.

PORTER, A. A.; VOTTA, L. G.; BASILI, V. R. Comparing detection methods for software requirements inspections: a replicated experiment. IEEE Transactions on Software Engineering, v. 21, n. 6, p. 563–575, jun. 1995.

Resource Description Framework (RDF) Model and Syntax Specification. Disponível em: <https://www.w3.org/TR/PR-rdf-syntax/>. Accessed on: 26 jul. 2023.

SANDAHL, Kristian, et al. An Extended Replication of an Experiment for Assessing Methods for Software Requirements Inspections. Empirical Software Engineering 3(4), 381-406, 1998.

Sharable Content Object Reference Model (SCORM®). Available at: <https://adlnet.gov/past-projects/scorm/#scorm-versions-and-resources>. Accessed on: 26 jul. 2023.

SOLARI, Martín; VEGAS, Sira. Classifying and analysing replication packages for software engineering experimentation. In: 7th International Conference on Product Focused Software Process Improvement (PROFES 2006)-Workshop Series in Empirical Software Engineering (WSESE). Amsterdam, Paises Bajos. 2006.

Understanding Metadata: What is Metadata, and What is it For?: A Primer | NISO website. Available at: <https://www.niso.org/publications/understanding-metadata-2017>. Accessed on: 29 jul. 2023.

WOHLIN, C. et al. Experimentation in Software Engineering. Boston, MA: Springer US, 2000.

WOHLIN, C. et al. Experimentation in Software Engineering. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.