

Adopting the Dublin Core Standard for Describing Open Scientific Data: The e-Quilt Prototype Experiment

Adriana Carla S. de Oliveira
University of Knoxville,
United States of America
adrianacarla.a@gmail.com

Guilherme Ataíde Dias
Federal University of
Paraíba, Brazil
guilherme@dc.ccsa.ufpb.br

Renata Lemos dos Anjos
Federal University of
Paraíba, Brazil
renatalemosdosanjos@gmail.com

Virgínia M. de Souza
Federal University of Paraíba, Brazil
virginiamirandadesouza@gmail.com

Pedro Luiz P. Corrêa
University of São Paulo, Brazil
pedro.correa@poli.usp.br

Keywords: open science; fourth paradigm; data life cycle; dublin core standard

1. The Fourth Paradigm and Open Data

The state of the art in scientific communication is centered on the fourth paradigm. Essentially it brings the open science, open scientific data and the management, sharing, aggregation, curation, preservation and scientific cooperation for the use and reuse of scientific research. We are in the era of intensive data. Hey apud Specht (2015) says:

This is one of the greatest motivations for the re-use of existing data for knowledge creation. With the advancement of technology in capturing and processing data, we have reached the fourth paradigm of data-intensive science and communication, where collaboration between different domain skill sets is required to successfully conduct meta-analysis. (Hey apud Specht, 2015).

Intensive data in the fourth paradigm reinforces the need to improve the skills and to adopt technologies, collaborative tools and methodologies in the context of open science.

Open-data has created an unprecedented opportunity with new challenges for ecosystem scientists. Skills in data management are essential to acquire, manage, publish, access and re-use data. These skills span many disciplines and require trans-disciplinary collaboration. (Specht et al., 2015, p.1)

The ongoing research relies on the data life cycle model and fourth paradigm. The data life cycle adopted for the stages development of the e-Quilt Prototype experiment is the Data Lifecycle developed by the DataONE initiative. This cycle is represented by 8 stages. Tenopir et al. (2011, p. 2) points out the importance of the model,

The collected data are processed through scientific data management and following the data lifecycle model. Different elements can be found in a dataset. For describing the dataset, it is necessary the adoption of metadata standards, follow the data lifecycle for its management and ensure their use and reuse in a long-term. In this way, “the data lifecycle cannot be considered independently from research lifecycle, as data are an indispensable element of scientific research.

The management of scientific open data is shown in Figure 1.

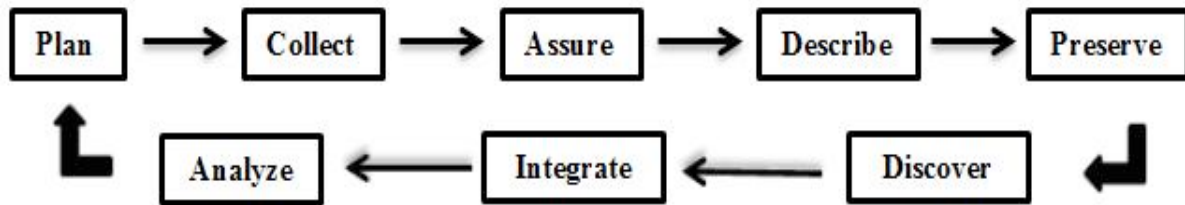


FIG. 1: Management of Scientific Open Data

This phase of the experiment is supported in the *Describe* stage. The data shared in the *e-Quilt Prototype* is the result of the research entitled, Epidemiological Survey on Oral Health, developed by the Department of Social Dentistry, UFPB, held in the cities of Caaporã and João Pessoa, Paraíba, Brazil, in the 2013-2015 period. Primary data collected were shared in the prototype and are being treated according the *Data Lifecycle*.

TABLE 1: Phases of experiment

PHASE 1	Deployment of the Dublin Core Standard to e-Quilt's Prototype Audio Resource	Metadata Description
PHASE 2	Use of the tool <i>Dublin Core Advanced Generation</i>	Automatic metadata code generation
PHASE 3	Audio Resource Metadata Adequacy Analysis	Resource Conformity to the Dublin Core Standard

To describe the metadata contained in the *e-Quilt Prototype* was used elements of the *Dublin Core* standard and the tool *Dublin Core Advanced Generation*.

1.1. Partial Results

The e-Quilt Prototype has the sufficient elements for metadata describing in conformity with international standards. It was verified that the sub-elements and the suggested resources in the Dublin Core standard are likely to be adopted by the metadata associated to the resource analyzed, as shown in Table 2.

TABLE 2: Results of application the audio resource in conformity with Dublin Core standard.

AUDIO RESOURCE – ATTRIBUTE CONFORMITY			
CONFORMITY	DC ELEMENT	CONFORMITY	DC ELEMENT
Yes	20	No	0
Partial	2	N/A	0

The *Identifier* and *Rights* elements associated to the resource presented partial compliance to the standard. The sub-elements DOI and ISBN associated to the *Identifier* element are not used. The analyzed resource is derived from the prototype that has no DOI and the ISBN does not apply to this resource, because it is applied to printed resources. The *Rights* element, presented partial compliance with the License sub-element and was described as unassigned. It was presented in the metadata that the audio resource is in accordance with the Brazilian Copyright Act (LDA - 9.610-1998). This analysis is guided by the adoption of a public license applicable to electronic publications on the international scenario.

The audio resource is derived from the main resource paper, both contained in the ambience of the e-Quilt Prototype. For the audio resource, it was found that it has considerable conformance

to the Dublin Core standard. As for the tool Dublin Core Advanced Generation tool adopted, it was observed that it has limitation concerning the automatic cleaning of characters (symbols, accents, etc.), which should be disposed manually when describing the metadata. Finally, it was analyzed that the description of metadata is a detailed process requiring the adoption of quality criteria and data validation.

References

- DataONE. Data Life Cycle model. Disponivel em: <<https://www.dataone.org>> Accessed: 2015 Mar 15.
- Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, version 1.1. (2014) Available: <http://dublincore.org/documents/dces/>. Accessed 2015 Mar 15.
- Specht, A., et al., Data management challenges in analysis and synthesis in the ecosystem sciences, *Sci Total Environ* (2015), <http://dx.doi.org/10.1016/j.scitotenv.2015.03.092> Accessed: 2015 Mar 05.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U. et al (June 2011). Data Sharing by Scientists: Practices and Perceptions, *PLoS ONE*. Volume 6, Issue 6. Available: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0021101> Accessed: 2015 Mar 06.