Chapter 8
Involving Data Creators in an Ontology–Based Design Process for Metadata Models

João Aguiar Castro
University of Porto, Portugal

Yulia Karimova
University of Porto, Portugal

Ricardo Carvalho Amorim
University of Porto, Portugal

João Rocha da Silva
University of Porto, Portugal

Rúbia Gattelli
University of Porto, Portugal

Cristina Ribeiro
INESC TEC/ DEI - University of Porto, Portugal

ABSTRACT

Research data are the cornerstone of science and their current fast rate of production is disquieting researchers. Adequate research data management strongly depends on accurate metadata records that capture the production context of the datasets, thus enabling data interpretation and reuse. This chapter reports on the authors' experience in the development of the metadata models, formalized as ontologies, for several research domains, involving members from small research teams in the overall process. This process is instantiated with four case studies: vehicle simulation; hydrogen production; biological oceanography and social sciences. The authors also present a data description workflow that includes a research data management platform, named Dendro, where researchers can prepare their datasets for further deposit in external data repositories.

INTRODUCTION

As the research environment is increasingly driven by data, research data management is gradually becoming a very important requirement for research projects. In the absence of proper management, expensive and irreplaceable research data may never realize their reuse potential; at the same time, their availability usually declines steadily as the publications age (Vines et al., 2014). While this is a problem for large-scale projects, it is even more prevalent in the context of research groups, or single researchers, in the long-tail of science (Heidorn, 2008), that often operate with very limited resources to ensure sustainability of their data.

To deal with this pressing issue, an increasing number of research funders are demanding research grant applicants to include data management plans in their project proposals - especially if public funds are required. These data management plans state, among other things, where and how the data will be deposited, preserved and kept accessible after the formal conclusion of the project. Major research funding providers are demanding such data management plans with recent calls for projects – examples include the European Commission under Horizon2020 (European Commission, 2013), and the National Science Foundation, in the US (National Science Foundation, 2011). Some publishers have also started to request data as supplementary materials to the submitted articles, under the assumption that their readers should be able to validate or replicate the presented results. Nature, for instance, requests authors to disclose research materials as a condition for the publishing of research papers. Another example is an Open Access publisher, PLOS ONE, that demands a full, unrestricted access to the original data for each of the submitted manuscripts. Following these trends data management is already an important concern for the scientific community.

The investment in research data management is important for many reasons: not only does it improve the chances of reproducibility and verifiability of the research results but can also prevent fraud. Another advantage of promoting data reuse relies on decreasing data duplication and the inherent research efforts to produce them. This allows researchers to directly focus their work in the project's specific goals, leaving more time to pursue an extensive validation or other research activities.

Research data management workflows involve both practical issues faced by the process stakeholders as well as technical ones. Sound technological solutions to support institutional repositories have been presented to reduce the technical issues, and we have seen great progress in that regard; solving the practical issues is, however, a challenge that is far from being settled, as it depends on fostering the interest of researchers to be active stakeholders in the data management workflow, more precisely in the description of their data. Data description assumes a critical nature in this workflow as it enables researchers with interest in a dataset to find and

Related Content

Multimedia Systems Development

Miloš Milovanovic, Miroslav Minovic, Velimir Štavljaninand Dušan Starcevic (2013). *Advancing Information Management through Semantic Web Concepts and Ontologies (pp. 86-104).*

www.irma-international.org/chapter/multimedia-systems-development/71850

Towards Web 3.0: A Unifying Architecture for Next Generation Web Applications

Tzanetos Pomonis, Dimitrios A. Koutsomitropoulos, Sotiris P. Christodoulouand Theodore S. Papatheodorou (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications (pp. 192-205).*

www.irma-international.org/chapter/towards-web-unifying-architecture-next/39171

Social Bookmarking and Web Search

Yusuke Yanbe, Adam Jatowt, Satoshi Nakamuraand Katsumi Tanaka (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications (pp. 242-259).*

www.irma-international.org/chapter/social-bookmarking-web-search/39174

Efficient String Matching Algorithm for Searching Large DNA and Binary Texts

Abdulrakeeb M. Al-Ssulami, Hassan Mathkourand Mohammed Amer Arafah (2017). *International Journal on Semantic Web and Information Systems (pp. 198-220).*

www.irma-international.org/article/efficient-string-matching-algorithm-for-searching-large-dna-and-binary-texts/189771

Semantic Web Technologies and Its Applications in Artificial Intelligence of Things

Shalini Roy, Harshit Gautamand D. Lakshmi (2024). *Semantic Web Technologies and Applications in Artificial Intelligence of Things (pp. 248-283).*

www.irma-international.org/chapter/semantic-web-technologies-and-its-applications-in-artificial-intelligence-of-things/347414