



Contribution ID: 176

Type: **Presentation**

The CODATA RDM Terminology: a community-focused approach to semantic interoperability

Thursday 16 October 2025 11:44 (11 minutes)

Introduction

As part of efforts to make research data more FAIR, semantic interoperability is important to consider. Standards, controlled vocabularies, and terminologies are well established types of FAIR-enabling resources that help us create interoperable systems and metadata. The CODATA Research Data Management Terminology (RDMT) is one such semantic resource that emerged from the former CASRAI Glossary to become a useful and usable, up-to-date reference tool for research data managers and other professionals involved in creating, managing and preserving research data. The CODATA RDMT is now published as a FAIR terminology through the Australian Research Data Commons (ARDC)'s Research Vocabularies Australia service.

This paper will discuss the purpose and value of the Terminology, and how we have taken a community-focused approach to its development to maximise the ability of the RDM community to contribute their expertise. We hope this will be of interest to those looking for a terminology for the RDM field for their own use, and/or those who are exploring approaches to terminology management and development.

Context

The CASRAI Glossary was intended as a practical reference for individuals and groups concerned with the improvement of research data management (RDM). In 2020, CASRAI asked CODATA to assume responsibility for the curation of this valued resource - a natural fit given CODATA's previous participation in the stewardship and development of the glossary, and our links with a heterogeneous range of task groups, working groups, national committees, and projects, giving us a rich network of expertise on which to draw.

The goal of the refreshed Terminology is to gather the key terms needed for a common understanding of the research data management domain. In this context, RDM refers to research data management practices covering the entire lifecycle of the data, from planning research to conducting it, and from backing up data as it is created and used to long-term preservation of data objects after the research investigation has concluded.

We realised that the real power of the resource was its ability to support meaning across different contexts, making it a terminology rather than a glossary, which resulted in the change of name. Then we recruited a Working Group to review the Terminology and contribute expertise from different relevant sectors, and established a rolling cycle of reviews with a fresh group of experts recruited for each review, to ensure diversity of input.

Method and approach

The RDMT is biennially reviewed and refreshed by an expert Working Group, which is responsible for creating a stable and sustainably governed standard terminology of community-accepted terms and definitions for concepts relevant to research data management, and keeping this terminology relevant by maintaining it as a 'living document' that is updated regularly. To those ends, the RDM Terminology Working Group uses a lightweight and pragmatic process to review the current Terminology and suggest any edits, additions and removals that are required to develop and improve the set of terms.

The Terminology is not an attempt to list every concept, tool and standard relevant to RDM; rather, it focuses on terms without easily found authoritative definitions elsewhere and offers an accessible definition in the context of contemporary RDM. Definitions are intended to be clear and unambiguous, and where possible, fit with common usage. We aim to produce definitions that are apposite across RDM activities of key stakeholders, including those working on research data management within the context of research, data management,

digital curation and preservation, research management, research policy, open data advocacy, computer science, information management, research administration, library, scholarly publishing, digital archiving and research funding roles. Some terms may have more than one definition, in which case the relevant context is specified.

Reflections and future directions

Our aim for the RDMT is to create and maintain the highest quality terminology possible within the bounds of our resources. To that end, we are working towards bringing definitions closer to the format specified in the ISO standard 704, for interoperability as well as for quality reasons. The approach to publication is also influenced by the principles laid out in the influential paper, “Ten simple rules for making a vocabulary FAIR”⁵, and is supported by our collaboration with ARDC to publish the Terminology in machine-actionable form.

We are keenly aware that the Terminology should serve as much of the global RDM community as possible. The 2025 review cycle involved participants based across thirteen countries. This is a welcome marked increase in geographical spread compared to the previous cycle which had representation from six countries. We are also encouraged by the interest shown in creating translations of the reference version. To date we have received enquiries about translations into other variants of English, variants of Chinese, French, Spanish and Portuguese.

We are also delighted that, in contrast to the broader trends within science⁵, we attract a high level of female expert participation: in our current Working Group, thirteen participants from nineteen are female. In the previous round, thirteen of seventeen participants were female. We are keen to celebrate and continue this success and do what we can to further improve diverse, cross-community participation.

Our paper will provide an overview of these various aspects of the development and management of the RDM Terminology, our approach to community review, and our plans for future development of this key resource to help improve semantic interoperability within the RDM and FAIR data communities.

Primary author: MOLLOY, L (CODATA)

Presenter: MOLLOY, L (CODATA)

Session Classification: Presentations Session 9: Empowering the global data community for impact, equity, and inclusion

Track Classification: SciDataCon Persistent Themes: Data Stewardship