

RESEARCH INFRASTRUCTURES IMPLEMENTING RDA OUTPUTS FOR MAPPING METADATA STANDARDS



Research Infrastructures and Virtual Research Environments in Europe implement RDA Metadata Standards Directory

There are several European Strategy Forum on Research Infrastructures (ESFRI) projects where metadata is a key component, in particular mapping in-use metadata standards in various domains to a canonical metadata scheme that is a conceptual and logical superset of the others (syntax and semantics). This forms the basis for defining interoperability converters to allow users of one ESFRI Research Infrastructure (RI) to use assets (data, software etc.) from another one or for a third party user to access and use assets from any of the ESFRI RIs involved.

The Metadata Standards Directory Working Group (MSDWG) and the Metadata Standards Catalog Working Group (MSCWG) outputs are used to discuss the mappings or potential mappings to produce the canonical superset metadata scheme and to inform about the metadata schemes used in various ESFRI Research Infrastructures (where they are recorded in the catalogue).

The Challenge

"The value inherent in shared research datasets can only be realized if peer researchers are able to identify, discover, contextualize, interpret and reuse them. They can only do this if the datasets are accompanied by metadata that describes, explains, and associates them with various other entities. When this metadata is missing or deficient, the dataset cannot be used to its full potential, and the scholarly endeavor is poorer as a result."

Says Keith G Jeffery Independent consultant working with various organisations.

There are several ways in which the metadata might be problematic. It might not exist at all. It might have to be deduced from unstructured text, a process which is prone to incompleteness and misinterpretation, and requires a level of human attention that is simply not scalable.

It might conform to an ad hoc or unsuitable standard, making it inconsistent with the metadata for peer datasets.

Such problems can be avoided if researchers endeavor to use existing standards wherever possible, to create local profiles instead of new specifications if the existing standards do not quite meet local needs, and to develop new standards only where there is a definite gap in provision.

In order to be able to do this, researchers need access to comprehensive knowledge of the metadata standards that are in use, both within their own field and generally across all fields.

RDA Output implemented

Metadata Standards Directory (MSDWG): enables discovery of metadata standards for documenting research data, regardless of academic discipline, and addresses issues related to coverage, ease of maintenance and sustainability

Answering community needs

The use of MSDWG/MSCWG outputs is applied to projects concerning Research Infrastructures and Virtual Research Environments in Europe. The organizations involved in this implementation process are ERCIM (www.ercim.eu) and NERC/BGS (<https://www.bgs.ac.uk>), while the key projects are EPOS (www.epos-ip.org) and ENVRIplus (<http://www.envriplus.eu>).

The main benefit for creators, editors, publishers, (re-)users of research data within these communities is the saved time in finding the relevant metadata 'standards'.

Why RDA

The RDA Metadata Standards Directory Working Group is supported by individuals and organizations involved in the development, implementation, and use of metadata for scientific data. The overriding goal is to develop a collaborative, open directory of metadata standards applicable to scientific data. A directory with similar aims had recently been developed independently by the UK Digital Curation Centre (DCC), so the group collaborated with the DCC on developing the directory further to achieve additional goals regarding coverage, ease of maintenance, and sustainability.

A second instance was designed in such a way as to simplify any future development effort, and indeed such development is being taken forward by the **Metadata Standards Catalog Working Group (MSCWG)**.

Find out more

Visit RDA @ rd-alliance.org
Email: enquiries@rd-alliance.org



The Implementation

The Metadata Standards Directory Working Group set out to develop a directory that would enable researchers, and those who support them, to discover metadata standards that would be appropriate for documenting their research data, regardless of their academic discipline. The aim is to get a consistent description of metadata 'standards' to have a basis for comparison and for planning how best to match and map research data. This way we could define an appropriate superset rich metadata standard to allow full interoperability across multiple research infrastructures and particularly across their data assets.

It is not a case of someone using the MSCWG catalogue to select a metadata standard and apply it to their use case – says Prof Keith G Jeffery - it is more the provision in a consistent form and at one location of information on metadata schemes useful for developing the interoperation converters."

The directory provided a consistent description of metadata 'standards' and useful ancillary information. The implementation process of using the outputs is ongoing; as it transforms into the catalogue this will become machine-manageable and thus more useful. The directory is by no means complete (and never will be since there are thousands of metadata 'standards'). However, the most common metadata 'standards' are there and the directory provides a good basis.

In the case of EPOS there are many metadata 'standards' in use in the various communities involved in different aspects of geoscience. However, some of them use INSPIRE/ISO19115 which is documented in the directory, and there are several variants of this standard used; being able to compare rapidly with the documented standard was useful.

In the case of ENVRIPlus the project is in the stage of discussion about the architecture and the use of catalogues – or one canonical superset catalogue – and the directory provides the basis for discussion of the optimal approach.

It provides a neutral documentation which is useful to drive consensus among passionate advocates of different approaches.

Finally, the VRE4EIC project has the objective of defining a reference architecture and component software for Virtual Research Environments. Again metadata interconversion to/from CERIF (Common European Research Information Format) – an EU recommendation to Member States – is a key aspect of the project, and the directory provides the basic information for metadata mapping and conversion.

Lesson Learnt

The major lesson learnt is that however good and complete the directory may be, there are always metadata 'standards' that are not registered. This is a continuous challenge, and requires some 'push' within the community to register the 'standards' they use so that others can benefit from their experience and effort.

A second lesson is that – while the directory records the standards and associated tools – there is a need to record interconversion tools which are critically important for constructing homogeneous access to heterogeneous research datasets (and associated software, publications etc.)

There is much more to be done.

Keith G Jeffery Consultants

Independent consultant, active member of RDA and co-chair of several groups working on issues related to metadata and Virtual Research Environments.



Prof Keith G Jeffery
CEng, CITP, FGS, FBICS, HFICS

Keith G Jeffery Consultants UK
(previously Director IT STFC Rutherford
Appleton Laboratory Chilton Didcot
Oxfordshire UK)

keith.jeffery@keithgjefferyconsultants.co.uk

+44 7768 446088