REPOSITORIES SUPPORT PROJECT

0845 257 6860

Metadata

support@rsp.ac.uk

Overview

Metadata is information about information. According to NISO it is “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”(1). Objects stored in an institutional repository are always accompanied by metadata records. This briefing paper introduces the different uses of metadata within an institutional repository and explores how such metadata can be defined.

Metadata in the repository

The primary function of metadata in an institutional repository is resource discovery. Resource discovery is enabled and enhanced by assigning relevant criteria to content items. This helps users identify resources, brings similar resources together, distinguishes dissimilar resources and gives location or access information. External systems also use metadata, harvesting it to provide, for example, interdisciplinary discovery services. Finally, maintaining good metadata also helps organise repository content and supports archiving and long-term preservation. Repository metadata has two main management contexts:

- Administrative: information collected from the author during the submission process, or input by the repository manager/administrator. These metadata fields are typically plain text administrative and bibliographic information input via a web form.
- Technical: Once the metadata is collected, it must be reliably stored and used to manage the resource. Certain technical protocols or standards, such as OAI-PMH(2), can be used to enable external services to access metadata records and facilitate resource discovery by a much broader audience.

It is important to consider these two contexts together. Strong communication between repository administrators and technical teams is important to ensure metadata input matches the harvesting requirements for external use.

Types of metadata

There are three main types of metadata:

- Descriptive: Facilitates resource discovery and identification; includes elements such as item titles, authors and keywords.
- Structural: Describes how items relate to one another; particularly important for recording versions of an item over time and for dealing with complex items such as chapters within a book or a conference paper that has been developed into a journal article.
- Administrative: Helps manage the resource itself; it includes rights management and preservation metadata that records how files were created and the rights associated with them.

Defining you metadata schema

Metadata schemas are sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. Most schemas will cover all three types of metadata identified above. Repository administrators will need to consider their metadata schemas at an early stage of repository implementation. Different types of information objects should be appropriately described. Local needs such as departmental and research structures, any planned integration with Library discovery platforms, and any local decisions needed about subject fields, should also be considered.

Repository managers will need to extend the schemas to cover new types of materials as the repository grows. In many instances the metadata schema which comes with the ‘out of the box’ installation of the chosen software will suffice to begin with. However, it should be carefully checked to ensure it meets and complies with basic requirements. Additional fields can be added should it be deemed necessary.

Metadata standards

A range of metadata standards have been developed by the repository and archives community that can guide those wishing to establish a schema. The most commonly used standard, and the one required for interoperability and harvesting via the OAI-PMH, is Dublin Core(3). Other metadata standards are available

for fulfilling more complex purposes, for example for collecting preservation metadata. These standards include METS(4), MODS(5), DIDL(6) and PREMIS(7).

Dublin Core

Dublin Core is a vocabulary of 15 properties for use in resource description that defines high level metadata elements such as date, creator, type and identifier. Dublin Core is widely used, simple and easily adopted as it is built into most repository software. It is also possible to qualify Dublin Core metadata to add further levels of detail to metadata, for example to distinguish between additional or translated titles, or define different types of dates or contributors.

Application Profiles

Application profiles are packages of metadata. Alternatively they can be defined as a declaration that specifies which metadata terms an organisation, information provider, or user community uses in its metadata schema. Several application profiles are being developed within the repository community including a scholarly works application profile – SWAP(8), an e-thesis application profile(9), and an images application profile(10).

CERIF(11) (Common European Research Information Format)

The CERIF data model is becoming increasingly important for repositories with the rise of Current Research Information Systems (CRIS) in UK Higher Education. Projects, funded by JISC and others, have been undertaken to explore the use of CERIF in repositories and to assist in recording impact(12). CERIF is currently managed by the euroCRIS(13) organisation and is not just a metadata standard but also a way to structure data within a system allowing you to develop complex, time-bound relationships between entities in the system. CERIF is essential in linking the disparate part of a CRIS together and for interoperability with external systems. This allows you to create the working life of a single researcher at a number of different institutions, or a snapshot of all the researchers working on X project at Y time. There are three central interlinking entities in CERIF, **person**, **project** and **organisation** and three central results, **patents**, **publications** and **products**. As an example: *at the point of writing '**Publication** P' has an author '**Person** A' who is a project leader for '**Project** X' as well as a member of department '**Organisation** Y', which is part of the university '**Organisation** Z' who employs '**Person** A'*.

Conclusion

Metadata is a vital element of any repository that supports day-to-day activities and keeps the repository usable. This paper has introduced a number of metadata standards and application profiles that have been developed to support metadata implementation and the development of appropriate schemas within institutional repositories. Wherever possible, schemas should be open and comply with widely recognised standards to facilitate interoperability and broader re-use of stored resources.

References and further information:

(1) NISO Understanding Metadata (http://www.niso.org/standards/resources/UnderstandingMetadata.pdf)
(2) OAI-PMH - Open Archives Initiative-Protocol for Metadata Harvesting (http://www.openarchives.org/)
(3) Dublin Core (http://www.dublincore.org/)
(4) METS – Metadata Encoding & Transmission Standard (http://www.loc.gov/standards/mets/)
(5) MODS - Metadata Object Description Schema (http://www.loc.gov/standards/mods/)
(6)DIDL - Digital Item Declaration Language (http://xml.coverpages.org/MPEG21-WG-11-N3971-200103.pdf)
(7) PREMIS - PREservation Metadata Implementation Strategies (http://www.oclc.org/research/projects/pmwg/)
(8) SWAP - Scholarly Works Application Profile (http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile)
(9) EThOS - UK Electronic Theses and Dissertations qualified Dublin Core (UKETD_DC) (http://ethostoolkit.rgu.ac.uk/wp-content/ethos-content/UKETD_DC.htm)
(10)Images Application Profile (http://www.ukoln.ac.uk/repositories/digirep/index/Images_Application_Profile)
(11) CERIF – Common European Research Information Format (http://www.eurocris.org/Index.php?page=featuresCERIF&t=1)
(12) Readiness for REF (R4R) (http://r4r.cerch.kcl.ac.uk/) and Measuring Impact under CERIF (MICE) (http://mice.cerch.kcl.ac.uk/)
(13)euroCRIS (http://www.eurocris.org)