Metadata Schema Registries

What is metadata?

Metadata is often described as “data about data”, information that typically describes what a document is, what it is about and where one may locate it (Zeng & Jin, 2008). The concept of metadata is not new; for every document, picture, video, or any type of file, metadata is generated by the computer or inputted by the user (e.g. File Name, Date Modified). What is new, is the potential that metadata provides in developing rich digital ‘libraries’.

Metadata Elements

Elements are the actual data fields, such as title, author, etc. The schema in which these elements are placed may be a flat structure, or may have more complex hierarchical structures, where an element ‘author’ may contain elements ‘name, Data of birth, Gender’.

Metadata Standards

In order to process metadata logically, it needs to be described in a standard way, hence metadata standards. The best known standard is the *Dublin Core* standard which provides an agreed set of core elements, 15 in fact: Title, Creator, Subject, Description, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage and Rights. Although this element set is clearly very basic, there are mechanisms for extending *Dublin Core* elements.

Brief Example

On the left, we have the element and on the right, we have the value.

Title	Bill and Ben meet the plant pot men
Creator	Smile E Face
Date	14/07/2009
Language	Flob-a--dob

Metadata Schema Registries

What is metadata? – Explains, very basically, what metadata is
Metadata Elements – Simple introduction to metadata elements
Metadata Standards – Explains metadata standards, focussing on Dublin Core
Brief example – Looks at a simple example of metadata
Why use a metadata schema registry – Outlines the reason behind a metadata schema registry
Architecture of a metadata schema registry – Brief explanation of the architecture
The Application Profile Complication – An insight into the world of Dublin Core

What is metadata?

Metadata is often described as “data about data”, information that typically describes what a document is, what it is about and where one may locate it (Zeng & Jin, 2008). The concept of metadata is not new; for every document, picture, video, or any type of file, metadata is generated by the computer or inputted by the user (e.g. File Name, Date Modified). What is new, is the potential that metadata provides in developing rich digital ‘libraries’.

Metadata Elements

Elements are the actual data fields, such as title, author, etc. The schema in which these elements are placed may be a flat structure, or may have more complex hierarchical structures, where an element ‘author’ may contain elements ‘name, Data of birth, Gender’.

Metadata Standards

In order to process metadata logically, it needs to be described in a standard way, hence metadata standards. The best known standard is the *Dublin Core* standard which provides an agreed set of core elements, 15 in fact: Title, Creator, Subject, Description, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage and Rights. Although this element set is clearly very basic, there are mechanisms for extending *Dublin Core* elements.

Brief Example

On the left, we have the element and on the right, we have the value.

Title	Bill and Ben meet the plant pot men
Creator	Smile E Face
Date	14/07/2009
Language	Flob-a--dob

Why use a metadata schema registry

The reason behind using a metadata schema is to avoid multiple schemas for similar or even identical data.

Zeng & Jin (p.274) provide the following general definition of metadata registries: “a metadata registry collects data regarding metadata schemas for reuse of existing metadata terms to achieve interoperability among metadata element sets. The basic components of a metadata registry may include identification of data models, elements, element sets, encoding schemas, application profiles, element usage information, and element cross>walks. The primary functions of metadata registries include registering, publishing and managing schemas and application profiles, as well as making the registry easily searchable within the registry. A registry also provides services for crosslinking and crosswalking among schemas and application profiles.”

Architecture of a Metadata Schema Registry

There are many possible ways of producing something that does this job. The IEMSR registry approach to represent the information in an RDF store, importing it from RDF files. Then it may be searched by means of SPARQL; a query language for RDF. This provides a ‘ready-made’ machine-to-machine interface which in theory can be replicated using any of a large set of open source or commercially available tools. Practically, the hard part of this problem is to decide what to represent and how it should be represented.

The Application Profile Complication

Originally, the application profile was a description of a specific form of usage of a schema, usually a more constricted set of guidelines than the original schema itself. The point of this was to customise schemas to specific needs and requirements.

However, nowadays, application profiles have a number of slightly more technical definitions; they may describe the extent to which one follows the guidelines of the original schema, or they may describe a ‘mix and match’ situation whereby metadata elements are selected from several metadata schemas to form one, optimised profile.

For new optimised application profiles to be formed, the registry must first contain the schemas and application profiles from which the elements are to be taken, which means that somebody somewhere first has to input every single schema/element/value to which that application profile refers.

References

Zeng, Marcia Lei, and Qin, Jain (2008). Metadata. Facet publishing. London, UK.

Why use a metadata schema registry

The reason behind using a metadata schema is to avoid multiple schemas for similar or even identical data.

Zeng & Jin (p.274) provide the following general definition of metadata registries: “a metadata registry collects data regarding metadata schemas for reuse of existing metadata terms to achieve interoperability among metadata element sets. The basic components of a metadata registry may include identification of data models, elements, element sets, encoding schemas, application profiles, element usage information, and element cross>walks. The primary functions of metadata registries include registering, publishing and managing schemas and application profiles, as well as making the registry easily searchable within the registry. A registry also provides services for crosslinking and crosswalking among schemas and application profiles.”

Architecture of a Metadata Schema Registry

There are many possible ways of producing something that does this job. The IEMSR registry approach to represent the information in an RDF store, importing it from RDF files. Then it may be searched by means of SPARQL; a query language for RDF. This provides a ‘ready-made’ machine-to-machine interface which in theory can be replicated using any of a large set of open source or commercially available tools. Practically, the hard part of this problem is to decide what to represent and how it should be represented.

The Application Profile Complication

Originally, the application profile was a description of a specific form of usage of a schema, usually a more constricted set of guidelines than the original schema itself. The point of this was to customise schemas to specific needs and requirements.

However, nowadays, application profiles have a number of slightly more technical definitions; they may describe the extent to which one follows the guidelines of the original schema, or they may describe a ‘mix and match’ situation whereby metadata elements are selected from several metadata schemas to form one, optimised profile.

For new optimised application profiles to be formed, the registry must first contain the schemas and application profiles from which the elements are to be taken, which means that somebody somewhere first has to input every single schema/element/value to which that application profile refers.

References

Zeng, Marcia Lei, and Qin, Jain (2008). Metadata. Facet publishing. London, UK.