

# Metadata Models – International Developments and Implementation

Heike Neuroth and Margo Bargheer

Göttingen State and University Library (SUB), Germany

## 1 Introduction into Metadata

Definitions of metadata often describe them as “data about other data”, sometimes refined through the expression “structured data about data”. This definition over-simplifies the facts, that metadata on one hand have been in use long before the digital age, be it in library catalogues or on inventory cards of museums, and that on the other hand the entity they represent does not necessarily need to be in form of bits and bytes. A highly generic definition for metadata therefore could be “(structured) information about (digital) objects”<sup>1</sup>. Transferred to the digital world “structured information” stands for “structured data”.

The term metadata does not solely refer to the representation of existing parts of reality such as data streams or objects, but is applicable as well to describe descriptions – metadata collections like subject gateways or digital libraries can be described in a metadata set as well. The unambiguous use of the term metadata therefore requires to regard the context as well as the levels of complexity to be described and the receivers of the information supplied through metadata. Metadata usually carry structured information like “Author”, “Title”, “Subject” etc, bits of information, that are semantically interconnected.

Instead of a fixed term or a standardised format metadata therefore should be understood as a form of language to exchange information on objects like books or digital resources as well as for purposes like archiving. Tom Baker describes Dublin Core<sup>2</sup>, one of the most common metadata formats, as “pidgin for Digital Tourists”<sup>3</sup>. Baker stresses the point that the structure of metadata forms a grammar, whose “basic pattern is easily grasped” and which “is well-suited to serve as an auxiliary language for digital libraries” (ibid.).

Metadata thus offer an elementary level of understanding, either between machines, human beings or in between the two of them.

---

<sup>1</sup> A more abstract definition could be “structured information on specific and describable sectors of reality”. “Specific and describable sectors of reality” could mean “a certain stream of data” as well as “an object”, “a document” or “a human being”. It should be emphasised, that metadata usually draw their meaning from the unambiguous relation to the reality they stand for. If there is no proof for the existence of the described, the description might be meaningless in the field of digital data.

<sup>2</sup> <http://www.dublincore.org/>

<sup>3</sup> <http://dlib.org/dlib/october00/baker/10baker.html>

### 1.1 Types and Function of Metadata

Who benefits from metadata, assumed they have been stringently applied? For the creators of digital objects (resources, documents, collections), be it a producer or a scientist, metadata offer the possibility to prepare a formalised description of their work and thus exercise control over it. For cataloguers metadata form an established tool for the description of particular types of resources. They enable implementors to set up reliable schema and information providers to share such reliable schema through cross-walks and harmonising processes in order to establish core sets, which are mentioned below in further detail. A major benefit of metadata is given to end users performing information retrieval, who can identify, locate and compare relevant resources and recognise the respective functional requirements.

According to their different statements on the nature of the described digital objects metadata can be divided into the following types:

- Content metadata/descriptive metadata  
refer to description of content or further bibliographic information, can be specified according to document type and/or subject. Technical details of real-life objects such as museum artefacts or books belong to this metadata type as well.
- Administrative metadata  
carry information for the distributed administration and the maintenance of archiving systems such as versioning of the metadata set or date stamps and signatures regarding metadata modifications.
- Structural Metadata  
allow the navigation in archiving systems by offering information on hierarchical levels (journal → article, monograph → chapter, artefact → detail).
- Technical Metadata  
carry information on the digital nature of the described resource/document such as size, format, resolution or colour.
- Preservation metadata  
inform about the durable preservation of digital objects and the storage/presentation format regarding migration and emulation.
- Terms and condition metadata  
allow the disclosure of copyrights, intellectual property and retrieval conditions such as payments or registration.
- Metadata about metadata collections  
allow users to cross-search distributed archiving systems (e.g. RSLP collection description<sup>4</sup>).
- Metadata about the use of metadata  
for example project/domain specific application profiles such as Renardus<sup>5</sup>, DC Libraries<sup>6</sup>, etc. Metadata of this type are usually stored in registries

<sup>4</sup> <http://www.ukoln.ac.uk/metadata/rslp/isadg/>

<sup>5</sup> <http://renardus.sub.uni-goettingen.de/renap/format.html>

<sup>6</sup> <http://www.dublincore.org/documents/2002/09/24/library-application-profile/>

carrying information on application profiles and element sets of the respective co-operation partners in distributed archiving systems.

Metadata feature certain aspects of functionality. Their service function allows structured access to different resources by offering possibilities like cross-searching, cross-browsing, result display, result ranking or result sorting. Their technical and administration functions allow long-term maintenance, metadata exchange and sharing as well as reliable archiving.

## 2 Application Profiles, Namespaces, and Registries

### 2.1 Application Profiles

To fully meet the potential of metadata as an elementary language for interoperability it is necessary to establish application profiles. Application profiles usually consist of several element sets<sup>7</sup>, but at least one of them. Application profiles can be used “as a way of making sense of differing relationship that implementors and namespace managers have towards metadata schema, and the different ways the use and develop schema.” (Heery/Patel 2000). Application profiles describe elements necessary for a certain implementation by specification of obligations such as “mandatory”, “strongly recommended” or “optional” in order to ensure that certain information such as “Title” and “Creator” will be supported by a specific project or domain specific implementation. Application profiles disclose which organising body or institution maintains which element set and provide guidelines and best practice for each element. Application profiles offer the possibility to shape domain-specific variations. Some examples:

- DC-Education<sup>8</sup>  
is based on the Dublin Core element set and includes elements from the IEEE Learning Object Metadata (LOM) element set. Target groups of education material for example thus can be specified on the metadata level already.
- DC-Government<sup>9</sup>  
includes all fifteen Dublin Core elements, supplemented with domain-specific elements refining information on rights (security classifications such as “top secret”), dates (reliable information on publishing dates), subjects, relation.
- DC-Libraries  
is based on the Dublin Core elements and includes elements of MODS (Metadata Description Object Schema), that allow the expression of roles for Creator/Contributor or the differentiation according to genre. DC Libraries support “Library of Congress Subject Headings” as an encoding scheme for Temporal Coverage.

<sup>7</sup> The term “Element set” should be used instead of the formerly term “namespace”, as “namespace” prevails as an XML-specific term defined and maintained by W3C.

<sup>8</sup> see DCMI Education Working Group,  
<http://www.dublincore.org/groups/education/>

<sup>9</sup> see DCMI Education Working Group,  
<http://www.dublincore.org/groups/government/>

- EULER AP<sup>10</sup>  
as a mathematics-specific application profile extends the DC elements by differentiating the role of the creator according to scientific publishing conditions or extending DC source type lists with EULER specific types.
- Renardus AP<sup>11</sup>  
Renardus as an interdisciplinary broker service offering sophisticated cross-browsing extends the DC elements by supporting several encoding schemas.

## 2.2 Element Sets (Formerly Called Namespaces)

Element sets describe a well defined set of metadata elements, according to domain or subject-specific requirements and to different implementations. Besides the pure graduation of information, element sets define semantics and syntax for each metadata element, serving as “a vocabulary that has been formally published, usually on the Web; it describes elements and qualifiers with natural language labels, definitions, and other relevant documentation.” (Baker 2000) Metadata generation should follow clear cataloguing rules. With Baker’s approach, understanding metadata element sets as a vocabulary, it is thus the underlying grammar to the vocabulary that conditions those cataloguing rules into a consistent semantics and syntax. This consistency is reached, when each element is treated as “a unique identifier formed by a name (e.g., Title)” (ibid.). Examples for element sets and their abbreviations are listed below:

- Dublin Core Metadata Element Set, Version 1.1<sup>12</sup>  
dc  
Dublin Core is one of the most common metadata formats maintained by the Dublin Core Metadata Initiative and it’s bodies, the respective working groups such as e.g. DCMI Libraries. The metadata element set can be refined in certain fields with Dublin Core Qualifiers (DCMI Metadata Terms<sup>13</sup>), abbreviated by “dc terms”.
- Agricultural Metadata Framework<sup>14</sup>  
fao  
stems from a common initiative by the FAO (Food and Agriculture Organisation of the United Nations) and OneWorld Europe<sup>15</sup>, which is based on Dublin Core, but specified according to the requirements of the agricultural field.
- IEEE Learning Object Metadata<sup>16</sup>  
ieee-lom

<sup>10</sup> <http://www.emis.de/projects/EULER/metadata.html>

<sup>11</sup> <http://renardus.sub.uni-goettingen.de/renap/renap.html>

<sup>12</sup> <http://www.dublincore.org/documents/dces/>

<sup>13</sup> <http://www.dublincore.org/documents/dcmi-terms/>

<sup>14</sup> <http://www.fao.org/agris/MagazineArchive/MetaData/TaskForceonDCMI.htm>

<sup>15</sup> <http://www.oneworld.org>

<sup>16</sup> IEEE (Institute of Electrical and Electronics Engineers, Inc.) Learning Technology Standards Committee (LTSC). Originating URL: <http://ltsc.ieee.org/wg12/>

is a subject specific element set for the field of education. Target groups of education material for example thus can be specified on the metadata level already.

- RSLP<sup>17</sup> Collection Description Metadata  
 rslpcld  
 refers to the special requirements the description of large collections pose.
- Metadata Object Description Schema (MODS)<sup>18</sup>  
 mods  
 “The Library of Congress’ Network Development and MARC Standards Office, with interested experts, has developed a schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. As an XML schema, MODS is intended to be able to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records. A mapping from Dublin Core metadata set to MODS schema is now available<sup>19</sup>.

### 2.3 Metadata Registry

Metadata element sets are mainly used for the mixing and matching of data. To ensure their reliability, interoperability and long-term maintenance, projects for setting up metadata registries have been started. They arose from the “recognition of the benefits of shared data dictionaries leading to the specification of a formal registration process in the standard ISO/IEC 11179.<sup>20</sup>” Metadata registries serve as reference tools for a wide range of complex data sets by promoting the re-use of already defined elements, disclosing data definitions on element sets used in local or subject-specific implementations, thus ensuring their authoritativeness. Other objectives are registries of controlled vocabularies within particular domains or developments of domain specific application profiles. SCHEMA and MetaForm are non-authoritative registries, which do not involve in the launching of definitions and standards:

- SCHEMAS<sup>21</sup> project, funded by the European Commission under the Fifth Framework Programme, “has provided a forum for metadata schema designers involved in projects under the IST Programme and national initiatives in Europe” (SCHEMAS Homepage). The CORES registry<sup>22</sup> will be developed to register application profiles and metadata element sets.
- MetaForm<sup>23</sup> (SUB), part of the German Meta-Lib<sup>24</sup> project, is a database with a special focus on the Dublin Core and its manifestations in various

<sup>17</sup> Research Support Libraries Programme, <http://www.rslp.ac.uk/AboutUs/>

<sup>18</sup> <http://www.loc.gov/standards/mods/>

<sup>19</sup> <http://www.loc.gov/standards/mods/dcsimple-mods.html>

<sup>20</sup> DELOS White Paper, 2003: Principles of Metadata Registries. Still to be published.

<sup>21</sup> <http://www.schemas-forum.org/>

<sup>22</sup> <http://www.cores-eu.net/>

<sup>23</sup> <http://www2.sub.uni-goettingen.de/metaform/index.html>

<sup>24</sup> Metadata Initiative of German Libraries,  
<http://www.dbi-berlin.de/projekte/einzproj/meta/meta00.htm>

implementations. MetaForm supports mapping processes between DC applications and other formats, crosswalks between the Dublin Core Metadata Element Set and its various dialects as well as crosscuts of a particular Dublin Core element (e.g DC.Creator) through different formats.

- The Dublin Core Metadata Initiative is the main authoritative body for defining and maintaining the Dublin Core element set. Terms and definitions, encoding schemes and qualifiers as well as controlled vocabulary systems used within the Dublin Core community need to be acknowledged by the usage board<sup>25</sup> to ensure Dublin Core's consistency and further development and are registered in a specific Dublin Core registry<sup>26</sup>.

Summarising the preceding paragraphs, the concept of application profiles and element sets, put down in metadata registries, follows certain objectives. They prevent the implementation of metadata schema based on in-house solutions, who are time-consuming and double effort and might work only locally. In this sense they increase the interoperability among different implementors dramatically and encourage cooperation between several partners resp. projects. The workability of metadata schema is ensured through the definition of unique identifiers for single elements as well as particular element sets or controlled vocabularies systems. These concepts ensure that bodies like management authorities feel responsible for the maintenance of metadata elements. Only durable and maintained concepts ensure interoperability and can serve as reliable interchange formats for cross-searching distributed services and heterogeneous environments.

### 3 Metadata Implementation

#### 3.1 Core Set of Metadata

The concept of metadata core sets is a prerequisite for an advanced service such as cross-searching. Cross-searching implies on the one hand, that searching over distributed and sometimes heterogeneous metadata collections via one single user-interface is technically supported, on the other hand, that searching results bring up comparable resources meeting the user's needs. To meet this objective, besides the technical development such as software architecture thorough preparatory work on the participating partner's metadata element sets is needed. This includes the following:

- analysis of semantics and syntax of each element
- investigation of all qualifiers in use. This refers to refinements as well as encoding schemes such as country-lists, classification systems or keywords, thesauri or controlled vocabularies

<sup>25</sup> The DC usage board consists of 7-11 members, who are knowledgeable about DC and actively work in the metadata community. Originating URL: <http://dublincore.org/usage/>

<sup>26</sup> see ongoing developments in the DCMI Registry Working Group, <http://www.dublincore.org/groups/registry/>

- analysis and harmonisation of cataloguing rules. This applies to core elements like “Title” or “Creator” as well as to “Description” and qualifiers such as keywords
- equalisation of rules on repeatability of each element
- analysis and harmonisation on obligation rules such as “mandatory”, “strongly recommended” and “optional”
- analysis of language qualifiers in use. It is highly recommended to investigate and harmonise language qualifiers for the fields “Title”, “Description” and “Subject”, especially in multilingual services or international co-operations that allow cross-searching over language boundaries

This preparatory work is usually conducted by detailed questionnaires and following discussions between the partners involved. In the end it should lead to the determination of a core set of metadata through identifying the minimum set of metadata elements that are needed to reasonably run the service and the maximum set of elements that each partner is able to support sufficiently. The definition of each element entering the core set is based on a “Format of Entry”.

### 3.2 Renardus Application Profile

Renardus aims to “provide a trusted source of selected, high quality Internet resources for those teaching, learning and researching in higher education in Europe. Renardus provides integrated search and browse access to records from individual participating subject gateway services across Europe.”<sup>27</sup> It will serve as an example to illustrate the concept of application profiles already mentioned. The application profile of Renardus is based on five element sets, encoded in XML/RDF.

- Dublin Core Metadata Element Set, Version 1.1: Reference Description [dc1.1]
- Dublin Core Qualifiers Element Set, which includes other DCMIS Elements and Dublin Core Qualifiers [dcterms]
- Dublin Core Type Vocabulary [dcmitype]
- Renardus Metadata Element Set [rmes]
- Renardus Metadata Element Set Qualifiers [rmesq]

The following examples show how different element sets have been implemented into the application profile. Cataloguing rules thus can be deduced from the application profile.

- Title and Title.Alternative  
title: dc1.1 (mandatory, not repeatable, language tag)  
title.alternative: dcterms (optional, repeatable, language tag)
- Creator  
dc1.1 (strongly recommended, repeatable)  
LastName, FirstName: rmesq (strongly recommended, repeatable)

---

<sup>27</sup> <http://www.renardus.org/>

- Description  
dc1.1 (mandatory in text version, repeatable, language tag)
- Country  
rmesq (strongly recommended, repeatable)

## 4 Long-Term Preservation

### 4.1 Preservation Metadata

Preservation Metadata inform about the durable preservation of digital objects and the storage/presentation format regarding migration and emulation. This includes the description of the complete “life cycle” of a digital object, especially when digital objects are not digital-born but results of transformation processes such as digitisation:

- provenance information like original format of the book
- date of digitisation
- technical information of the digitisation process
- presentation and storage format (supporting migration and emulation)
- RMS (rights-management-system), supplying information on terms and conditions of access, copyright and intellectual property
- etc.

There are a lot of projects and initiatives which are working on preservation metadata. The following list is not complete but provides an overview:

- Preservation Metadata for Digital Objects: A Review of the State of the Art, A White Paper by the OCLC/RLG Working Group on Preservation Metadata (January 31, 2001),  
[http://www.oclc.org/research/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/research/pmwg/presmeta_wp.pdf)
- Preservation Metadata and the OAIS Information Model, A Metadata Framework to Support the Preservation of Digital Objects, A Report by The OCLC/RLG Working Group on Preservation Metadata (June 2002),  
[http://www.oclc.org/research/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/pmwg/pm_framework.pdf)
- A Recommendation for Preservation Description Information, A Report by The OCLC/RLG Working Group on Preservation Metadata (April 2002)  
[http://www.oclc.org/research/pmwg/pres\\_desc\\_info.pdf](http://www.oclc.org/research/pmwg/pres_desc_info.pdf)
- A Recommendation for Content Information, A Report by The OCLC/RLG Working Group on Preservation Metadata (October 2001),  
<http://www.oclc.org/research/pmwg/contentinformation.pdf>
- Metadata Encoding and Transmission Standard (METS),  
<http://www.loc.gov/standards/mets/>
- National Library of Australia – Preservation Metadata for Digital Collections, Exposure Draft,  
<http://www.nla.gov.au/preserve/pmeta.html>

- Metadata for Preservation – CEDARS Project Document AIW01 (August 1998),  
<http://www.ukoln.ac.uk/metadata/cedars/AIW01.html>
- Cedars Guide To : Preservation Metadata (March 2002),  
<http://www.leeds.ac.uk/cedars/guideto/metadata/>
- NEDLIB: Metadata for long term-preservation (July 2000),  
<http://www.kb.nl/coop/nedlib/results/D4.2/D4.2.htm>
- Rebecca Guenther: The joint work of the OCLC-RLG Preservation Metadata Working Group, RLG Open Forum at ALA June 16, 2002,  
<http://www.rlg.org/longterm/forum02/guenther.html>

#### 4.2 OAIS (Open Archival Information System Reference Model)<sup>28</sup>

The OAIS is a conceptual framework for an archival system important for the long-term preservation of digital objects, stemming from the work of the space data community. Since its beginnings in 1997 the framework has gained international recognition due to the common effort of RLG<sup>29</sup>, OCLC, and many members of their respective organisations in shaping the reference model and adapting it for the use in libraries, archives and research repositories. Especially archive designers and maintainers can benefit from a reliable framework such as OAIS, which provides common concepts and terminology. The OAIS has been implemented already into the NEDLIB project<sup>30</sup>, run by the Koninklijke Bibliotheek, Den Haag. The participation of international organisations ensures the maintenance and further development of the framework.

#### 4.3 Open Issues

Although OAIS provides a sophisticated framework, several issues still need to be developed in the future. This especially refers to distributed archiving systems such as EMANI, the electronic mathematical archiving network initiative<sup>31</sup>. Projects like EMANI work on concepts, how to handle the granularity of digital objects (such as journals/article or monograph/chapter) in archiving systems. Other questions to be solved refer to the development of a minimum set of preservation metadata in distributed services. Automatic generation of metadata and preservation metadata standards, best practices and clear guidelines display fields for further engagement as well.

### 5 Outlook

The questions thrown open by this article refer to certain related topics. Further developments from there thus promise to be fruitful for those questions. Schema

<sup>28</sup> <http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>

<sup>29</sup> <http://www.rlg.org/rlg.html>

<sup>30</sup> <http://www.kb.nl/coop/nedlib/>

<sup>31</sup> <http://www.emani.org/>

issues for example will be touched by progresses in XML/RDF research, architecture issues will benefit from progresses on the field of de-central discovery systems such as Z39.50<sup>32</sup>, LDAP<sup>33</sup> or OAI<sup>34</sup>. Architectural issues will influence at least the administrative metadata already mentioned. Research on formats will refer to presentation and storage format, which is important for migration and emulation of data. It may be expected, that additional metadata are needed to cover the aspect of format sufficiently. Business models, which concern metadata for rights, permission of use, payments and registrations are important as well. Trends and developments in this field should be observed carefully, as they might call for modifications in the respective metadata elements. Metadata are indispensable for the efficient search across multiple collections by supporting interoperability and crosswalks. A prerequisite for such services are reliable registries. Thorough application and use of metadata supports the documentation and maintenance of interrelationships within repositories. With descriptive and technical metadata for example, digitisation processes can be traced. For the dissemination of digitised contents, metadata are a major tool for resource discovery. They allow the documentation of multiple versions of digital objects, be it updated versions, different formats or translations. Metadata not only describe these versions but allow connections and links between the respective objects. Rights and reproduction information are stored safely in metadata. Ambitious applications in the Web such as the large-scale preservation of the Cultural Heritage or developments of the Semantic Web would be unthinkable without metadata.

## References

1. Baker, Thomas: A Grammar of Dublin Core. D-Lib Magazine October 2000. Originating URL: <http://dlib.org/dlib/october00/baker/10baker.html>
2. Heery, Rachel, Patel, Manjula: Application profiles: mixing and matching metadata schemas. Ariadne Issue 25, 24-Sep-2000. Originating URL: <http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>

---

<sup>32</sup> <http://www.loc.gov/z3950/agency/>

<sup>33</sup> see e.g. <http://www.openldap.org/>

<sup>34</sup> <http://www.openarchives.org/>