A Metadata Application Profile for the German Virtual Library

Carola Wessel
Project META-LIB
Heike Neuroth
Project Renardus
Goettingen State and University Library
D-37070 Goettingen, Germany
wessel@mail.sub.uni-goettingen.de

Abstract

A number of German virtual subject libraries cooperate to establish "The German Virtual Library." To enable cross-searching and cross-browsing over their distributed resources, they decided to develop a metadata application profile, based on the DCMES. The answers to a questionnaire were evaluated to figure out which elements were supported by the individual partners. This paper describes the approach taken as well as the discussion on the use of certain elements and introduces the draft of the application profile.

Keywords: *Metadata Application Profile, German Virtual Libraries, DCMES, Interoperability*

1. German virtual subject libraries

This paper deals with the development of a metadata application profile for the German Virtual Library, a cooperation of a number of virtual subject libraries. The concept of these virtual subject libraries is based on the system of supra-regional literature supply, which is a joint project of numerous scientific libraries throughout Germany, supported by the German Research Foundation (Deutsche Forschungsgemeinschaft). In combination with The German Library (Die Deutsche Bibliothek) and The Collection of German Prints (Sammlung Deutscher Drucke), this system of supra-regional literature supply can be seen as the equivalent of a national library. The participating libraries are each responsible for a specific special subject collection, meaning that they are assigned to the task of providing access to all relevant literature in their field. The special subject collections are structured into subject oriented ones (e.g. French Language and Literature, Economics, Physics) and regional collections (e.g. Africa South of the Sahara). All in all there are 121 special subject collections which can be found in more than 40 libraries all over Germany.

The fast change of the information technology directly affects the literature and information supply. There are numerous modifications in the publication and information system, e.g. more and more publications appear in electronic or digitized form. This leads to new challenges for the system of the supra-regional literature supply. In 1998 the German Research Foundation decided to extend the scope of the special subject libraries to include digital and electronic resources. Thus the concept of virtual subject libraries was developed.

Following this concept, a virtual subject library is more than only a central access point to electronically available documents of a certain subject field. Instead, it is an entrance portal to basically all relevant information and documents in this field, regardless of its form as printed, electronic or digitized material and regardless of the physical residence of the document. Only the relevance of the content is important. Ordering and delivery services shall be improved so that the researcher can use all the material he or she needs at his or her desk.

These tasks are not to be solved by the individual research and technical libraries on their own. Instead, they cooperate in one virtual library, The German Virtual Library [1]. In addition to searching possibilities independent of the form of a document, here the user can search over a number of different subject areas. To enable access to the collections of all the virtual subject libraries via one portal, it is necessary to integrate the different technical systems, formats and documentary languages.

In the first phase of this project, the 15 participating virtual subject libraries are:
Anglo-American History and Anglo-American Literature (SUB Goettingen)
Business Economics (USB Cologne)
Modern art (SLUB Dresden)
History (BSB Munich)
Medicine (ZB Med Cologne)

https://doi.org/10.23106/dcmi.952106701

Netherlands culture area (ULB Muenster)
Pharmacy (UB Braunschweig)
Political sciences and peace research (SUB Hamburg)
Psychology (SULB Saarbruecken)
Social sciences (IZ Soz Bonn)
Technics (UB/TIB Hanover)
Veterinary medicine (BTiHo Hanover)
The Middle East, including Northern Africa (ULB Halle)
Political economics (ZBW Kiel)

A number of additional partners will join this group soon. As can been seen from the topics included in this list, the range of fields covered in The German Virtual Library is quite diverse.

2. The development of the metadata core set

To enable interoperability and search features across all participating libraries, it was decided to rely on metadata. A working group on metadata was established who agreed on the use of the Dublin Core Metadata Element Set as the basis for their own application profile. META-LIB[2], the German initiative for the use of metadata in libraries, was asked to develop recommendations for a specific core set of metadata as an application profile for the German Virtual Library.

The partners of META-LIB are The German Library[3] and Goettingen State and University Library[4]. What will be presented in the following are the results of the work done by Hans Becker, Christel Hengel, Heike Neuroth, Berthold Weiss and Carola Wessel.

The basis for the application profile we develop will be the Dublin Core Metadata Element Set (DCMES)[5] and DCMES with qualifiers[6]. Since we realized early on that we would need additional elements, it was decided to develop a domain-specific Virtual Library Metadata Element Set with qualifiers, VMESq (so far, no new simple elements have been added and no simple VMES has been developed). Therefore, our application profile will consist of metadata drawn from these namespaces. In combining them, we follow the example of the DC education working group which developed its own DC Education profile[7].

All partners are expected to map their own sets of metadata to this common set and thus enable cross-searching and cross-browsing. Individual partners may add further elements to this set if needed.

In addition, this set will serve as a model for future virtual libraries in Germany who are expected to develop their own set of metadata according to this profile.

META-LIB compiled a questionnaire to find out more about the usage of metadata in German virtual libraries[8]. It consisted of three parts: Part one asked for general information on the virtual library such as addresses, subject areas covered and resource types. Part two covered codes, formats, standards, classifications, thesauri etc. in use by the partners, further divided in printed, digitized and online resources as well as CD-ROM. The third part dealt with the metadata concept. Based on the DCMES, we asked for the usage of each element and its qualifiers (element refinements and encoding schemes). In addition, we inquired if the elements were mandatory, recommended or optional and whether they could be repeated or not. A further question was whether the partners used the specific element according to the definition provided by the DCMI. Since many answered "yes" while later on giving differing information, this question was neglected in our evaluation. This paper will concentrate on the third part.

In the development and the evaluation of this questionnaire, we relied on the experiences made in the EU-project Renardus, a collaborative project that aims to improve academic users' access to a range of existing Internet-based information services across Europe[9].

We received 14 completed questionnaires[10]. While analyzing the answers, it became clear that some questions had been misunderstood and others had not been phrased precise enough. So we added a second round where we sent out further inquiries.

The answers were evaluated and summarized in a quite extensive paper which listed how many partners supported certain elements, including the codes and standards used for each element. Based on this survey, we formulated a first draft for the use of metadata in which we named the elements and qualifiers that were supported by many partners and should be included in the application profile because they would allow for cross-searchability. Semantics and structure for each element were defined. We also included unresolved issues.

This paper was discussed at two meetings with the partners. It became obvious that we had a number of differing opinions on certain aspects. We decided to add a Guide of Best Practice which contains further explanations of our recommendation to ensure that all partners use the elements in the same way. Also, examples may be added to describe the use of an element. Based on these answers and discussions, we are developing a Metadata Application Profile for German Virtual Libraries.

While we were discussing our profile, the DC Library Working Group started to develop a library-specific application profile[11]. The draft will be reviewed at the meeting in Tokio, and we will follow the discussion closely.

The following results are still preliminary since we will need to further discuss them with the partners before coming to definite conclusions.

https://doi.org/10.23106/dcmi.952106701

In the following, we will introduce the answers we received as well as the discussions that evolved and then present our recommendations for a metadata core set for a German Virtual Library. In addition to naming the elements and qualifiers, we will note their obligation (mandatory, recommended or optional) and repeatability. During the cause of the evolution, we realized that it would be helpful to specify the definition of mandatory: Some elements are mandatory in any case (mandatory 1), others only if it is possible to detect the content (mandatory 2). Although all elements are optional and repeatable in the DCMES, we still consider our use of the elements as according to Dublin Core even when we prescribe obligations or deny repeatability.

Concerning the DC elements that are supported by the partners, the following list emerged: All 14 libraries use the elements Title, Creator, Subject, Description, Publisher, Date, Identifier and Language. Format and Coverage are supported by 13 partners, Relation by twelve, Type and Rights by eleven, Contributor by eight and finally Source by six libraries. This extensive conformance is due to the fact that all partners already used the DCMES as the basis for their individual metadata set.

The answers become much more diverse when one looks at the statements regarding to obligation and repeatability and at the qualifiers in use.

Our format of entry for the data model looks like this:

Name	Name of metadata field
Namespace	Choice of namespace: DCMES version 1.1. DCMES with Qualifiers (2000-07-11) Virtual Library Metadata Element Set (VMES) VMES with Qualifiers
DC Refinement(s)	These qualifiers make the meaning of an element narrower or more specific. A refined element shares the meaning of the unqualified element, but with a more restricted scope.
DC Encoding Scheme(s)	These qualifiers identify schemes that aid in the interpretation of an element value. These schemes include controlled vocabularies and formal notations or parsing rules. A value expressed using an encoding scheme will thus be a token selected from a controlled vocabulary or a string formatted in accordance with a formal notation. If an encoding scheme is not understood by a client or agent, the value may still be useful to a human reader. The definitive description of an encoding scheme for qualifiers must be clearly identified and available for public use.
VLib Refinement(s)	See above.
VLib Encoding Scheme(s)	See above.
Form of obligation	There are four grades of obligation: Mandatory 1: This element has to have a value in any case. Mandatory 2: This element has to be supported when possible. Recommended. Optional.
Repeatability	The metadata field is repeatable: Yes or no.
Language Qualifier	To give information about the language of the content of the metadata field. Yes or no.
DC Definition	Dublin Core definition of the metadata element.
DC Comment	Dublin Core comments to its metadata element.
VLib Definition	Virtual Library definition of the metadata element.
VLib Comment	Virtual Library comments to its metadata element.
Guide	Best practice guide.
Example	Example for the use of the metadata element.

This format is modelled on similar Dublin Core and Renardus formats[12].

Since it would be too voluminous to describe each element according to this model, the results will be summarized.

GENERAL ISSUES

Since some issues relate to more than one element, these will be named here: We decided to use ISO 3166 (2-letter-code)[13] for countries, ISO 639-2[14] for languages, ISO 8601[15] for dates, MIME types for formats and UNICODE[16].

We will introduce a new field that gives information on the origin of the content of the field, e.g. if a description was taken from the source or added by a cataloger. It will be used as an attribute with standardized vocabulary.

TITLE

In the case of Title, the answers were very much in unison: All partners support this element, it is mandatory for all and only two partners allow to repeat it. The element Title.Alternative is used by

https://doi.org/10.23106/dcmi.952106701

nine partners, it is mandatory in one case and repeatable in all cases.

It is recommended to use Title in the meaning of Main Title and put all other forms (translations, additions) in Title.Alternative.

Our application profile for Title looks like this:

DC.Title: mandatory 1, not repeatable

DC.Title.Alternative: recommended, repeatable

A language qualifier should be used for both entries.

SUBJECT

This element is supported by all partners, it is always mandatory and repeatable. The encoding scheme used most frequently is the DDC (Dewey Decimal Classification)[17].

We had some discussion on the use of a common encoding scheme which is necessary to enable cross-browsing. Some partners favoured the BK (Basisklassifikation, a classification that was developed in the Netherlands and is common in libraries who work with the PICA system) because it is not as granular as the DDC and used by most of them for printed material. Since German libraries want to be able to cooperate with international partners and since we can rely on positive experiences in the project Renardus, we decided to use the DDC. In addition, tailored to the German users, we will recommend the use of the SWD (SchlagWortnormDatei, a German thesaurus)[18]. This scheme will be taken from the Virtual Library Metadata Element Set (VMES). The use of specific classifications and thesauri for certain subject areas is recommended. These may be taken from the DCq schemes or added to the VMESq schemes. We will establish a registry with unequivocal descriptions of the individual schemes.

Furthermore, we decided that the element Subject should only be used with a scheme and not in its simple form. To allow the use of free keywords, we add an extra scheme.

Although it was seen as desirable by some to establish the qualifier Keywords and Classification, we at last decided against it in line with the DCMES.

So we recommend the use of:

DCq.Subject Scheme=DDC: recommended, repeatable

DCq.Subject Scheme=[special classification or thesaurus] recommended, repeatable

VMESq.Subject Scheme=SWD: recommended, repeatable

VMESq.Subject Scheme=[special classification or thesaurus] recommended, repeatable

VMESq.Subject Scheme=FreeKeyword: optional, repeatable, with language qualifier

DESCRIPTION

Description is supported by all partners, mostly mandatory (eleven times) and repeatable (nine times). The refinement Description.Abstract is used ten times (mandatory: seven, repeatable: nine) and Description.TableOfContents four times (not mandatory, repeatable: three).

Additional qualifiers like review may be considered. It should be indicated whether an abstract was copied from the original resource or written by the cataloguer. Quoting of the URL is an optional addition.

The application profile looks like this:

DC.Description: mandatory 2 (meaning that one of the fields has to be used, either the simple form or a qualifier), repeatable

DCq.Description.Abstract: optional, repeatable

DCq.Description.TableOfContents: optional, repeatable

All elements should contain a language qualifier.

CREATOR, CONTRIBUTOR, PUBLISHER

Since the discussion on the use of these elements is still going on in the DC Libraries WG as well as in the DC Agents WG, we will follow this discussion closely. If they decide to replace these three elements with a single element (whether it will be called Agents or Contributor), we will adopt this approach. Therefore, we will only list some of our results and discussions for these elements.

A topic that needs further discussion is the linking of names to authority files. All partners agreed that it would be very helpful to recommend certain schemes, but we still habe to decide which ones. In the German context, the PND (PersonenNamenDatei, German standard for individual names)[19] for personal names and the GKD (Gemeinsame KörperschaftsDatei, German standard for corporations)[20] for corporate names could be applied, but it is also important to use international authority files.

We agreed that the syntax should follow the order FamilyName, GivenName.

CREATOR

Creator is supported by all partners, is mandatory in six cases and always repeatable.

PUBLISHER

All partners use this element, it is mandatory nine times and repeatable eight times.

CONTRIBUTOR

This element is used by only eight libraries, it is mandatory in one case and repeatable in six cases.

DATE

Date is supported by all partners, is mandatory in eight cases and twice repeatable. The most frequently used refinements are Date.Modified (seven entries) and Date.Created (six entries). As scheme ten partners use W3C-DTF[21], and we decided to make the use of this scheme mandatory.

It is still to be decided whether we need more refinements. There was also some discussion whether a date can be repeatable. Since it might happen that we have different URIs linking to different versions

https://doi.org/10.23106/dcmi.952106701

of an document, we decided to make it repeatable. Since a date is not always available, its obligation is mandatory 2.

Concerning the qualifiers, we are still discussing whether Date.Modified is useful because this date would need to be checked every day to be correct. New qualifiers are Date.Submitted, Date.Accepted and Date.Archived. These are taken from the DissOnline project[22] and seem especially valuable for dissertations and journal articles.

In addition to the refinements mentioned, we use the administrative metadata element Date.Metadata. LastModified. We may also add an element Date.Metadata.Created to note when the resource was added to the database.

We recommend:

DCq.Date Scheme=W3C-DTF: mandatory 2, repeatable

DCq.Date.Created: recommended, not repeatable

DCq.Date.Issued: recommended, repeatable

DCq.Date.Modified: recommended, repeatable

VMESq.Date.Submitted: optional, not repeatable

VMESq.Date.Accepted: optional, not repeatable

VMESq.Date.Archived: optional, not repeatable

VMESq.Date.Metadata.LastModified: recommended, repeatable

TYPE

Type is used by all partners, is mandatory ten times and always repeatable.

It is not yet clear if we will develop our own type list or adopt an existing list. We will follow the discussions on DC Type and other lists. For now, we recommend the use of the DCMIType Vocabulary[23].

The profile contains:

DCq.Type Scheme=DCMITypeVocabulary: mandatory 1, repeatable

FORMAT

All but one of the partners support this element, it is mandatory in ten cases and repeatable in eight cases. As refinements five partners use Extent and ten use Medium.

We still need to discuss obligation and repeatability.

The profile contains:

DCq.Format.Extent: optional, repeatable

DCq.Format.Medium: recommended, repeatable

DCq.Format Scheme=IMT[24]: recommended, repeatable

IDENTIFIER

This element is used by all partners, it is mandatory in eleven cases and repeatable in three cases. The refinement URI is supported by eleven libraries, it is mandatory for all and repeatable for three.

We decided that Identifier has to be used with a URI. Possible schemes would be URL, URN, DOI, ISBN or ISSN. A unequivocal, non-repeatable identifier should point to an unequivocal main title. Also, we consider the establishment of new refinements like Archive or Mirror to enable an explicit assignment of a URI to a specific document. The use of these refinements would be especially helpful with repeated identifiers.

Questions arise concerning synonym URLs (more than one URL for one web site) and documents with multiple sites.

For now, we recommend:

DCq.Identifier Scheme=URI: mandatory 1, not repeatable

VMESq.Identifier.Mirror Scheme=URL: optional, repeatable

VMESq.Identifier.Archive Scheme=URL: optional, repeatable

SOURCE

The element Source is used six times, is mandatory twice and repeatable four times.

Instead of Source, Relation.IsPartOf or Relation.IsFormatOf could be used. The application of Source is helpful for digitized or filmed items. This topic has to be discussed further.

Recommendation:

DC.Source: optional, repeatable

LANGUAGE

All partners support this element, it is always mandatory and repeatable in twelve cases. The encoding scheme ISO 639-2 is used twelve times.

The application profile will look like this:

DCq.Language: mandatory 2, repeatable, Scheme=ISO 639-2

RELATION

This element was used by twelve partners, is mandatory once and repeatable seven times. Of the refinements IsPartOf is used most frequently (ten times), followed by IsVersionOf, HasVersion and HasPart (three times each).

Relation.IsFormatOf should only be used in cases where the content of the resource has been changed. Otherwise VMESq.Identifier.Mirror or …Archive should be used. The meaning of “has”-relations need further discussion.

Recommendation:

DC.Relation: optional, repeatable

DCq.Relation.IsPartOf: recommended, repeatable

All other refinements are optional.

COVERAGE

This element is used 13 times, it is mandatory three times and repeatable five times. The refinement Spatial was mentioned twice, the refinement Temporal ten times. As schemes DCMI Period[25] (five partner) and W3C-DTF (four partner) are used.

A specific problem with Coverage is the distinction between this element and Subject. We

https://doi.org/10.23106/dcmi.952106701

suggest that Coverage is mainly used for the description of temporal periods.

Coverage should only be used with qualifiers.

We recommend:

DCq.Coverage.Spatial: recommended, repeatable, Scheme=TGN[26]

DCq.Coverage.Temporal: recommended, not repeatable, Scheme=DCMI Period[27]

DCq.Coverage.Temporal: recommended, not repeatable, Scheme=W3C-DTF

RIGHTS

Rights is supported by eleven partners, it is mandatory for ten and repeatable for ten.

We are still discussing the use of VMES Qualifier to specify the usage of rights since we are convinced that this element will become more and more important.

The application profile will contain:

DC.Rights: mandatory 2, repeatable

So far, we have not agreed upon adding new elements to our application profile. However, a number of elements were listed as desirable:

Country; Evaluation; refinements PersonalName and CorporateName for Creator, Contributor and Publisher.

We are not considering administrative metadata, technical metadata or metadata for archiving purposes at this point. Since there are some projects dealing with these aspects, we will discuss these topics with them.

3. The application profile

The short version of the complete Application Profile of German Virtual Libraries looks like this (Please note that this is still a draft.):

OTHER ELEMENTS

Element	Refinement	Scheme	Obligation	Repeatability
DC.Title			M 1	no
	DCq.Title.Alternative		R	yes
DCq.Subject		DCq=DDC	R	yes
		DCq=[special class./thesaurus]	R	yes
		VMESq=SWD	R	yes
		VMESq=[spec.class./thesaurus]	R	yes
		VMESq=FreeKeyword	R	yes
DC.Description			M 2	yes
	DCq.Description.Abstract		O	yes
	DCq.Description.ToC		O	yes
DC.Creator		To be discussed	M 2	yes
DC.Publisher		To be discussed	M 2	yes
DC.Contributor		To be discussed	R	yes
DCq.Date		DCq=W3C-DTF	M 2	yes
	DCq.Date.Created		R	no
	DCq.Date.Issued		R	yes
	DCq.Date.Modified		R	yes
	VMESq.Date.Submitted		O	no
	VMESq.Date.Accepted		O	no
	VMESq.Date.Archived		O	yes
	VMESq.Date.Metadata. LastModified		R	yes
DCq.Type		DCq=DCT 1	M 1	yes
DCq.Format	DCq.Format.Extent		O	yes
	DCq.Format.Medium		R	yes
		DCq=IMT	R	yes
DCq.Identifier		DCq=URI	M 1	yes
	VMESq.Identifier.Mirror		O	yes
	VMESq.Identifier.Archive		O	yes
DC.Source			O	yes
DC.Language			M 2	yes
		DCq=ISO 639-2	R	yes
DC.Relation			O	yes
	Relation.IsPartOf		R	yes
	Relation….		O	yes

https://doi.org/10.23106/dcmi.952106701

DC.Coverage			O	yes
	DCq.Coverage.Spatial	DCq=TGN	R	yes
	DCq.Coverage.Temporal	DCq=DCMI Period	R	no
	DCq.Coverage.Temporal	DCq=W3C-DTF	R	no
DC.Rights			M 2	yes

[1] <http://www.virtuellefachbibliothek.de>
[2] <http://www.sub.uni-goettingen.de/nojava_home.htm>, <http://webdoc.gwdg.de/edoc/aw/bfp/preprint/2001/wessel.pdf>
[3] <http://www.ddb.de>
[4] <http://www.sub.uni-goettingen.de>
[5] <http://www.dublincore.org/ documents/dces/>
[6] <http://www.dublincore.org/documents/dcmes-qualifiers/>
[7] <http://www.ischool.washington.edu/sasutton/dc-ed/Dc-ac/DC-Education_Report.html>
[8] <http://www.ddb.de/etc/ FormularTeil1.html> (in German)
[9] <http://www.renardus.org/>
[10] For the results, see <http://www2.sub.uni-goettingen.de/metacore/>
[11] <http://www.dublincore.org/documents/2001/08/08/library-application-profile/>
[12] The definitions are taken from the DCMES (see notes 4 and 5), for the Renardus model see <http://renardus.sub.uni-goettingen.de/renap/format.html>
[13] <http://www.din.de/gremien/nas/nabd/iso3166ma/codlstp1/index.html>
[14] <http://www.loc.gov/standards/iso639-2/>, this code can be easily mapped to the ISO 639-1, see <http://www.ifla.org/IV/ifla65/papers/099-155e.htm>
[15] <http://www.uic.edu/year2000/datefmt.html>
[16] <http://www.unicode.org/>
[17] <http://www.oclc.org/dewey/about/index.htm>
[18] <http://www.ddb.de/professionell/swd.htm>
[19] <http://www.ddb.de/professionell/pnd.htm>
[20] <http://www.ddb.de/professionell/gkd.htm>
[21] <http://www.w3.org/TR/NOTE-datetime>
[22] <http://www.dissonline.de/>
[23] <http://dublincore.org/documents/dcmi-type-vocabulary/>
[24] <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>
[25] <http://dublincore.org/documents/dcmi-period/>
[26] Getty Thesaurus of Geographic Names, <http://www.getty.edu/research/tools/vocabulary/tgn/>
[27] <http://dublincore.org/documents/dcmi-period/>

https://doi.org/10.23106/dcmi.952106701