EBU core metadata set for Radio archives

December 2001

European Broadcasting Union
Case postale 45
Ancienne Route 17A
CH-1218 Grand-Saconnex
Geneva, Switzerland
techreview@ebu.ch

EBU·UER

Contents

Scope

This EBU document defines a simple set of metadata which is adapted for use in radio archives, but which is aligned both with the main metadata standards of the broadcasting industry (EBU/SMPTE/AES) and with the Dublin Core metadata (the general approach used by libraries and archives, as well as the worldwide web [1]). An appendix to this document gives some background information on Dublin Core.

The metadata set defined in this document contains fifteen individual metadata elements, with their qualifications, in accordance with the Dublin Core [2], and recommended usages.

The appendix also gives information on the relationship between the metadata set defined in this document and the overall standardization of metadata by the EBU.

As far as possible, this document refers to controlled vocabularies (Authority Lists) for use in the metadata elements. However, the EBU has developed some specialist vocabularies [1] for use in broadcasting. These are used for the elements: role, type and format.

The EBU intends to develop a system for the expression of the metadata for use in actual implementations.

1. Currently these lists are under development by the EBU but preliminary lists, which may be extended, are available on the EBU website (see the **References** section at the end of this document).

Metadata for Radio Archives

Elements

The following set of elements is based upon the Dublin Core set, extended with qualifiers to comply with the special needs for audiovisual resources within the area of broadcasting material and other audiovisual media [3] [4] [5].

Element 1:	**Title**
Name:	Title
Definition:	A name given to the resource.
DC Comment:	Typically, a title will be a name by which the resource is formally known.
EBU Comment:	The title is tied to the archived item: for a series – use the series title; for a programme – a programme title; for an item – an item title. To differentiate between a series title and programme title when these are identical, recommended best practice is to use a date along with the programme title. For example, "News" is a series title; "News 2000.11.12" is a programme title. Titles are recorded as they appear.
Element refinement:	Alternative
Comment:	Any form of the title used as a substitute or alternative to the formal title of the resource.

Element 2:	**Creator**
Name:	Creator
Definition:	An entity primarily responsible for making the content of the resource.
DC Comment:	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
EBU Comment:	It is recommended that names are written in the following order: surname, first name; however they can also be written according to local practice.
Element Refinement:	Role
Element Encoding scheme:	EBU Reference Data Table: Roles in broadcasting [6]
EBU Comment:	The element refinement Role is added by EBU and not part of standard DCMES. The content of the qualifier Role must be taken from a controlled list of authorized roles. The Role Authority List [6] is recommended, but can be extended to cover special local needs.

Element 3:	**Subject**
Name:	Subject and Keywords
Definition:	The topic of the content of the resource.
DC Comment:	Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

EBU Comment:	Persons as subjects are also placed here. Genre of the content is placed under element Type.
Element Encoding schemes:	Library of Congress Subject Heading (LCSH) [7] Library of Congress Classification (LCC) [8] Medical Subject Headings (MeSH) [9] Dewey Decimal Classification (DDC) [10] Dansk decimalklassedeling 5.utgave (DK5) [11] Klassifikasjonssystem för svenska bibliotek (SAB) [12] Universal Decimal Classification (UDC) [13] Norske emneord [14]

Element 4: Description

Name:	Description
Definition:	An account of the content of the resource.
DC Comment:	Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.
Element Refinement:	Table of Content
Definition:	A list of subunits of the content of the resource.
Element Refinement:	Abstract
Definition:	A summary of the content of the resource.
EBU Comment:	For a Radio or television programme a running order can be used as description. EBU recommends not using the element refinements here, but instead just write a description of the content under the element itself.

Element 5: Publisher

Name:	Publisher
Definition:	An entity responsible for making the resource available
DC Comment:	Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
EBU Comment:	The publisher given here is the entity that has made the resource available in digital form.

Element 6: Contributor

Name:	Contributor
Definition:	An entity responsible for making contributions to the content of the resource.
DC Comment:	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
EBU Comment:	If in doubt whether an entity is a creator or contributor use the element contributor. It is recommended that names are written inverted; surname, first name, but they can also be written according to local practice.
Element Refinement	Role
Element Encoding scheme:	EBU Reference Data Tables: Roles in broadcasting [6]
EBU Comment:	The element refinement Role is added by EBU and not part of standard DCMES. The content of the qualifier Role must be taken from a controlled list of authorized roles. It is recommended that roles are taken from the EBU Reference Data Table [6], but this list can be extended to cover special local needs.

Element 7:	**Date**
Name:	Date
Definition:	A date associated with an event in the life cycle of the resource.
DC Comment:	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 and follows the YYYY-MM-DD format.
Element Encoding scheme:	DCMI: Period Encoding Scheme [15] W3C-Date and Time Format (DTF) [16]
Element Refinement:	Issued
DC Comment:	Date of formal issuance (e.g. publication) of the resource.
EBU Comment:	It is recommended to use this element to show the date the content was issued, i.e. the broadcasting date of a radio programme. It is recommended best practice to use the element both for recordings that are "born-digital" and recordings that are digitized. The date for digitizing is placed in Date.Digitized.
Element Refinement:	Created
EBU Comment:	It is recommended to use this element to show the date the content was created. It is recommended best practice to use the element both for recordings that are "born-digital" and recordings that are digitized.
Element Refinement:	Digitized
EBU Comment:	The element refinement Digitized is added by EBU and not part of standard DCMES. The date an analogue recording is digitized is placed here.

Element 8:	**Type**
Name:	Resource Type
Definition:	The nature or genre of the content of the resource.
DC Comment:	Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the working draft list of Dublin Core Types [17]). To describe the physical or digital manifestation of the resource, use the FORMAT element.
EBU Comment:	If it is necessary to give more than one type repeat the type element.
Element Encoding Schemes:	DCMI: Type vocabulary [17] EBU Reference Data Table: Type of resource [18] RDS: PTY display terms [19]

Element 9:	**Format**
Name:	Format
Definition:	The physical or digital manifestation of the resource.
DC Comment:	Typically, Format may include the media-type or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [20] defining computer media formats).
Element Refinement:	Extent
Definition:	The size or duration of the resource
EBU Comment:	Recommended best practise is to use Extent to give the duration of a recording. The duration should be written in the form HHMMSS.
Element Refinement:	Medium
Definition:	The material or physical carrier of the resource.

EBU Comment: The content of the qualifier follows a list of authorized Media.

Element Encoding schemes: Internet Media Type (IMT) [20]
EBU Reference Data Table: Storage media [21]

Element 10: Identifier

Name: Resource Identifier

Definition: An unambiguous reference to the resource within a given context.

DC Comment: Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Example formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

EBU Comment: Recommended best practice is to use the SMPTE Unique Material Identifier (UMID) for broadcasting material resources and other appropriate identifiers such International Standard Recording Code (ISRC) or International Standard Audiovisual Number (ISAN) for editorial content.

Element Encoding Schemes: SMPTE Unique Material Identifier (UMID) [22]
Uniform Resource Identifier (URI) [23]
International Standard Recording Code (ISRC) [24]
International Standard Audiovisual Number (ISAN) [25]

Element 11: Source

Name: Source

Definition: A Reference to a resource from which the present resource is derived.

DC Comment: The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.

EBU comment: The Recommended best practice is to use a unique identifier to identify the physical source that has been used to create the digital resource. In the case of a digitized analogue recording, it is the recording used for digitization which is the source. For commercial recordings the label and number is normally given here. For talking books the ISBN and/or label and number for the talking book (not the original book) is given here. If no label or number is available, the title and/or the statement of responsibility etc. of the digitized recording is recorded here. For a digitized radio programme the production number is normally given here.

Element 12: Language

Name: Language

Definition: A language of the intellectual content of the resource.

DC Comment: Recommended best practice for the values of the Language element is defined by RFC 1766 [26], which includes a two-letter Language Code (taken from the ISO Standard 639 [27]), followed optionally, by a two-letter Country Code (taken from the ISO Standard 3166 [28]). For example, 'en' for English, 'fr' for French, or 'en-uk' for English used in the United Kingdom.

EBU Comment: Recommend best practise is to use the ISO-639-2 - 3-letter codes.

Element Encoding scheme: ISO-639-2 [27]
W3C RFC 1766 [26]

Element 13:	**Relation**
Name:	Relation
Definition:	A reference to a related resource.
DC Comment:	Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.
EBU Comment:	Relation is used to show the relation in content to another resource. For example, "IsPartOf" is used to show the relation between a part of a radio programme and the whole programme, or between a track and a record album. If it is more relevant, information other than an identifier, e.g. a Title, can be given to show the relation to another resource.
Element Refinement:	IsVersionOf HasVersion IsReplacedBy Replaces IsRequiredBy Requires IsPartOf HasPart IsReferencedBy References IsFormatOf HasFormat
Element Encoding schemes:	URI [23]

Element 14:	**Coverage**
Name:	Coverage
Definition:	The extent or scope of the content of the resource.
DC Comment:	Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [29]) and that, where appropriate, named places or time periods be used in preference to numeric identifiers such as sets of coordinates or date ranges.
EBU Comment:	Coverage is used to show various time and place aspects of the subject of the content. For example the geographical origin of folk music is placed here.
Element Refinement:	Spatial
Definition:	Spatial characteristics of the content of the resource.
Element Encoding Scheme:	ISO 3166 [28] Getty Thesaurus of Geographical Names [29] DCMI: Point Encoding scheme [30] DCMI: Box Encoding Scheme [31]
Element Refinement:	Temporal
Definition:	Temporal characteristics of the content of the resource.
Element Encoding Scheme:	DCMI: Period Encoding scheme [15] W3C-Date and Time Format [16]

Element 15:	**Rights**
Name:	Rights Management
Definition:	Information about rights held in and over the resource.

DC Comment: Typically, a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.

EBU Comment: By “Rights” we here mean the rights to the programme (sound file etc.) as a whole. For rights covering parts of the file (music, poetry etc included in the file) keep these in the institutions locally. Here register for instance NRK, BBC or the production company responsible for the programme or the record company that owns the rights to a phonogram. Here one can use an URL to point to places with more information on rights.

Appendix A
Relationships with other metadata schemes

Background to Dublin Core and SMPTE metadata and sets

What is metadata?

Metadata literally means "data about data". Any catalogue – card or online – contains metadata. But today, the term is increasingly applied by information professionals to the value-added information that they create to arrange, describe, track and otherwise enhance access to information objects.

Why metadata?

Metadata is used to describe, in a standardized way, the information necessary to locate a document. The aim is to provide a minimum set of metadata that is understood and used by everyone. A standard set of metadata will:

- Provide a standard way to describe network-accessible material;
- Enable more precise queries to be made;
- Help search engines present "hits" grouped by subject rather than as a random mix.

Why Dublin Core?

Dublin, Ohio, USA is the home of OCLC (Online Computer Library Centre). The Dublin Core 15 Element Set was proposed and published as DC version 1.0 in December 1996 by the Dublin Core Metadata community.

The Dublin Core Metadata Element Set (DCMES) grew out of a recognized need for improved resource discovery of web resources. Initially it focused on the requirement of simplicity: "ordinary" users should be able to formulate descriptive records based on a relatively simple scheme. But over the years there has been a movement to use DCMES for more complex and specialized resource description tasks and, correspondingly, to develop mechanisms for incorporating such complexity within the basic element set. This work is called *qualified Dublin Core*.

There is a consensus within library, archive and web-technology communities that Dublin Core is a suitable general approach for standardization of metadata. It has obtained increasing support since it was consolidated in 1996 and it is obvious that it has many qualities:

- It is a relatively simple format which can be extended, without limit, with local fields;
- It has international support;
- It has proved helpful to users in finding things;
- It is widely recognized and supported;
- It can be used directly in websites and as records in a database because of the way it is structured;
- It is maintained in a stable environment;
- Its continuing development seems assured.

And, it is proving to be hospitable to a wide range of disciplines and domains, including sound recordings and moving images.

The Core elements

The elements are listed in the order they were developed by the Dublin Core Metadata Initiative (DCMI), but there are other useful ways to group them. In the following table, you can see that some elements relate to the content of the item, some to the item as intellectual property, and still others to the particular instantiation, or version, of the item.

Table 1: Grouping of Dublin Core elements

Content	Intellectual Property	Instantiation or Version
Coverage	Contributor	Date
Description	Creator	Format
Type	Publisher	Identifier
Relation	Rights	Language
Source		
Subject		
Title		

Definitions of Qualifiers:

Element refinement:

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. A client that does not understand a specific element refinement term should be able to ignore the qualifier and treat the metadata value as if it were an unqualified (broader) element. The definitions of element refinement terms for qualifiers must be publicly available.

Element encoding scheme:

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 (e.g., a term from a classification system or set of subject headings) or a string formatted in accordance with a formal notation (e.g., “2000-01-01” as the standard expression of a date). If a client or agent does not understand an encoding scheme, 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.

Relationship of this document to overall EBU metadata standardization

This recommendation covers the essential metadata that radio archives would associate with the exchange of radio material. It has a particular value for the discovery (search and retrieval) of content in a large archive. It also has value for supporting common, EBU-wide, access to archive holdings.

EBU metadata elements and attributes

It is anticipated that the individual metadata elements defined in this EBU document will be fully compatible with other EBU metadata standardization. This scheme is under development by the EBU project P/META [32]. When the full EBU metadata standard is published, the elements in this present document will be capable of being formally identified (mapped) in terms of the units of any more general EBU standard.

EBU metadata sets

The EBU draft metadata scheme provides a structure, called a Set, to group useful metadata elements. The Set construction allows a formal definition of the mapping from the 15 Dublin Core elements to elements or sets of elements drawn from the SMPTE Metadata Dictionary [33].

Relationship of this document to SMPTE metadata

Metadata elements

The SMPTE metadata dictionary is one of a number of metadata tools developed as a result of the need for standardization – originally identified by an EBU/SMPTE Task Force Report [34]. As well as a dictionary of metadata elements, SMPTE also define:

- Registries to provide additional control for SMPTE metadata elements [33];
- Structured use of metadata elements through the definition of metadata sets [35].

The metadata elements described in this EBU document are intended to fully align with elements of the SMPTE metadata dictionary or with formally defined sets of such elements. (See the section below.) For this reason, the elements in the present standard on metadata for radio archives have been mapped to elements of the EBU P/META scheme, in the anticipation that these will in turn be harmonized with the SMPTE dictionary.

Metadata sets

The SMPTE has defined a set structure for metadata elements. The EBU intends that the content of sets defined in the EBU metadata scheme will be harmonized with the contents of equivalent sets registered by the SMPTE.

Expression of the metadata

This standard does not specify how the actual metadata is held or transported. Work is in progress to define transport mechanisms for metadata, both when embedded with material or transported separately.

References

[1] **Dublin Core Metadata Element Set, Version 1.1: Reference Description**.
http://dublincore.org/documents/dces/

[2] **Dublin Core Qualifiers**.
http://dublincore.org/documents/dcmes-qualifiers/

[3] **Using Dublin Core**.
http://dublincore.org/documents/usageguide/

[4] Internet RFC 2413: **Dublin Core Metadata for Resource Discovery**.
Internet Engineering Task Force.
http://www.ietf.org/rfc/rfc2413.txt

[5] **Katalogiseringsregler och formatinstruktion för en svensk Dublin Core-tillämpning**
(av Sten Hedberg).
http://www.kb.se/Nvb/Katalog/DC/format1.htm *(In Swedish)*

[6] **EBU Reference Data Tables: Roles in broadcasting**.
http://www.ebu.ch/tech_info_roles.html

[7] **LCSH - Library of Congress Subject Heading**.
http://lcweb.loc.gov/cds/lcsh.html

[8] **LCC - Library of Congress Classification**.
http://lcweb.loc.gov/catdir/cpso/lcco/lcco.html

[9] **MeSH - Medical Subject Headings**.
http://www.nlm.nih.gov/mesh/meshhome.html

[10] **DDC - Dewey Decimal Classification**.
http://www.oclc.org/dewey/index.htm

[11] DK5 : decimalklassedeling / udgivet af Dansk BiblioteksCenter. - [Ballerup] :
Dansk BiblioteksCenter, 2000.

[12] SAB – Klassifikationssystem för svenska bibliotek. - 7., omarb. uppl. / utarbetad på uppdrag av SAB:s kommitté för katalogisering och klassifikation; under redaktion av Anders Noaksson; i samarbete med Klassifikationsgruppen inom SAB:s kommitté för katalogisering och klassifikation.
Lund : Bibliotekstjänst, 1997.

[13] **UDC - Universal Decimal Classification**.
http://www.udcc.org/

[14] Norske emneord.
Oslo : Biblioteksentralen, 1996.

[15] **Dublin Core Metadata Initiative – Period Encoding Scheme**.
http://dublincore.org/documents/dcmi-period/

[16] W3C Note: **Date and Time Formats**.
Worldwide Web Consortium.
http://www.w3.org/TR/NOTE-datetime

[17] **Dublin Core Metadata Initiative – Type Vocabulary**.
http://dublincore.org/documents/dcmi-type-vocabulary/

[18] **EBU Reference Data Tables: Type of resource**
http://www.ebu.ch/tech_info_media.html

[19] **Radio Data System (RDS): PTY display terms in English, Dutch, French, German, Norwegian, Portuguese, Spanish and Swedish – Issue: 2.0 (1998)**
RDS forum – the association of EDS users, Geneva.
http://www.rds.org.uk/rds98/pty_languages.htm

[20] **Internet Media Types**.
http://www.isi.edu/in-notes/iana/assignments/media-types/media-types

[21] **EBU Reference Data Tables: Storage media**.
http://www.ebu.ch/tech_info_store.html

[22] SMPTE STANDARD 330 for Television: **Unique Material Identifier (UMID)**.
The Society of Motion Picture and Television Engineers, 1999.

[23] RFC 2396: **Uniform Resource Identifiers (URI): Generic Syntax**.
Internet Engineering Task Force.
http://www.ietf.org/rfc/rfc2396.txt

[24] ISO Final Draft International Standard 3901: **ISRC – International Standard Recording Code**.
http://www.nlc-bnc.ca/iso/tc46sc9/3901.htm

[25] ISO Draft International Standard 15706: **ISAN – International Standard Audiovisual Number**.
http://www.nlc-bnc.ca/iso/tc46sc9/15706.htm

[26] RFC 1766: **Tags for the Identification of Languages**
Internet Engineering Task Force.
http://www.ietf.org/rfc/rfc1766.txt

[27] ISO 639-2: **Codes for the representation of names of languages – Part 2: Alpha-3 code (Registration Authority)**.
http://lcweb.loc.gov/standards/iso639-2/

[28] ISO 3166: **Codes for the representation of names of countries**.
http://www.din.de/gremien/nas/nabd/iso3166ma/index.html *(from 2002, this site will be hosted by the ISO – see* http://www.iso.org/iso/en/catsupport/maralist.html*)*.

[29] **Getty Thesaurus of Geographic Names**
http://www.getty.edu/research/tools/vocabulary/tgn/index.html

[30] **Dublin Core Metadata Initiative – Point Encoding Scheme**.
http://dublincore.org/documents/dcmi-point/

[31] **Dublin Core Metadata Initiative – Box Encoding Scheme**.
http://dublincore.org/documents/dcmi-box/

[32] **EBU P-META Metadata Scheme** *(To be published)*.

[33] **SMPTE Metadata Dictionary** *(as specified in SMPTE RP210a)*
http://www.smpte-ra.org/mdd/index.html

[34] Final Report of the EBU/SMPTE Task Force
http://www.ebu.ch/pmc_es_tf.html#rep

[35] SMPTE Standard 336M-2001 for Television: **Data Encoding Protocol using Key-Length-Value**.
The Society of Motion Picture and Television Engineers.