

Radio Archive Metadata

Richard Wright

BBC Information & Archives

Marit Grimstad

NRK Digital Radio Archives

Radio archivists and engineers across Europe have agreed on a simple set of terms for describing archive content. This set agrees with the standard widely used in conventional archives, libraries, publishing and web production – and by the Audio Engineering Society.

This article describes what was agreed and why, and how it fits in with other metadata work in broadcasting.

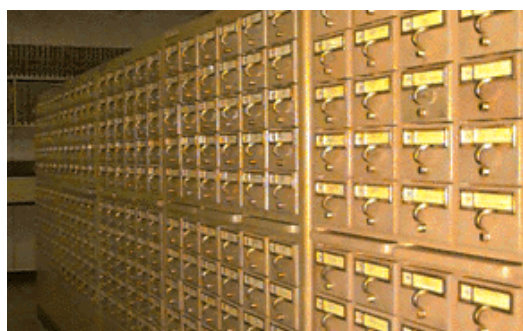
Introduction

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

An example of an electronic (web) document which happens to be about metadata – and the metadata describing the document – are shown in *Appendix A*.

Metadata is used to describe, in a standardized way, the minimum set of information that is necessary to locate a document – or, in the case of broadcasters, to locate a programme.

- metadata provides a standard way to describe network-accessible material;
- metadata enables you to make more precise queries;
- metadata helps the search engines to present hits that are grouped by subject rather than a random mix.



Metadata – for finding a needle in a haystack; a card catalogue is the metadata needed to find the needle. The only problem is: haystacks don't have catalogues!

What is – and isn't – metadata?

The data held in a database could be names and addresses, or appointment times, or parts lists, or many many other kinds of information – but it will usually consist of just text. The term “metadata” when applied to databases is really the naming (labelling) of the data elements. **There is a lot of text in databases which is NOT metadata.**

In the world of documents, the contents are text (which may be on a computer and might be searchable), but there is also a requirement to label the documents in order to find them in large collections. Hence the use of catalogues and the need for document identification – the metadata. **There is a lot of text in documents which is NOT metadata.**

But in the world of broadcasting, audio and video signals are the chief interest. For these signals to be managed, stored and retrieved, they also need labels – just like documents. So text is used to label broadcasting signals, leading to the notion that audio and video are data, and if it's text, it must be metadata.

This view is too simple, and leads to problems. In particular, in metadata standardization, it leads to the effort to standardize all forms of text, including all text data in all databases, under the assumption that everything in broadcasting that isn't audio or video but, instead, is text, must therefore be metadata. **There is a lot of text in broadcasting which is NOT metadata.**

The situation isn't clear-cut, because information has many uses. A script, contract or cast list is a document for the purposes of our document archive, and has a little bit of associated identifying metadata (usually a programme number). But the information in a script, contract or cast list is useful for describing the associated audio and video signals – and can even be useful for finding those signals. For example, cast lists if held in a database would allow the retrieval of all programmes in our archive where a certain actor had a role. In this case, the cast list information is used for one of the main functions of metadata – for finding things. So the cast list is text, it is used to find programmes – it must be metadata (it walks like a duck, it sounds like a duck). However, the platypus shown on the right has webbed feet and a bill, but it isn't a duck. **Not all text is metadata.**



It would be preferable to say that all text is potentially useful because of the associations of the information it represents. It would be preferable not to say that all text and all textual data is metadata, because that leads in the standardization world to defining a problem that is too large to solve – the standardization of all text and text data used in broadcasting.

That large problem is avoided by restricting the standardization process to the essential information needed to describe and retrieve programmes, and related elements. So instead of worrying about all forms of text and text data, the key to efficient progress is to concentrate on the core information.

The recommendation of P/FRA

The EBU working group *Future Radio Archives* (P/FRA) has completed the task of defining a minimum set of metadata for retrieving material (video as well as audio) from broadcast archives, and for exchanging this material with other broadcasters and other archives. The result of this work is Tech 3293: *EBU Core Metadata Set for Radio Archives* [1]. The work of the group benefited enormously from the work already done in the Scandinavian countries by SAM, the Scandinavian Audiovisual Metadata group. SAM already had an approach and a working document when P/FRA first met – the task of P/FRA was to establish whether the approach had general consent, and to work out whether the approach was compatible with overall EBU metadata activity.

The SAM document defined 15 items of core metadata (shown in *Table 1* below) which were not invented by SAM but are an existing standard – Dublin Core – which is already widely supported.

Why Dublin Core?

Background (origin), dissemination, recognition

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 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 the 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, which began with the community of “web resources” (and includes library and archive communities), that Dublin Core is a suitable general approach for the standardization of metadata. Dublin Core is now a US NISO standard (Z39.85) and ratification by ISO (TC 46) and CEN is in progress. 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 that 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

In Tech 3293, the core elements are listed in the order in which they were developed by the Dublin Core Metadata Initiative (DCMI) [2], but there are other useful ways to group them. In *Table 1*, you can see that some elements relate to the content of the item, some to the item as intellectual property, 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		

To make these elements specific, unambiguous and helpful in broadcasting, Tech 3293 gives three further sorts of information:

- 1) an interpretation of the element for the purposes of broadcasting;
- 2) where we thought it necessary, we break down (refine) the element to allow greater detail;
- 3) we provide controlled text (lists; encoding schemes) for certain elements, to allow (rather, to force) broadcasters to use a common terminology where that approach is possible.

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 to overall EBU metadata standardization

Document Tech 3293 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 Tech 3293 will be fully compatible with other EBU metadata standardization, under development by the EBU project group, P/META [3].

When the full EBU metadata standard is published, the elements in Tech 3293 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 [4].

Relationship of EBU Tech 3293 to SMPTE metadata

Metadata elements

The SMPTE metadata dictionary is the one of a number of metadata tools developed as a result of the need for standardization originally identified by an EBU/SMPTE Task Force as reported in [5]. As well as a dictionary of metadata elements, the SMPTE also defines:

- a) registries to provide additional control for SMPTE metadata elements [4];
- b) structured use of metadata elements through the definition of metadata sets [6].

The metadata elements described in Tech 3293 are intended to fully align with elements of the SMPTE metadata dictionary or with formally defined sets of such elements.

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.

Relationship of EBU Tech 3293 to AES metadata

The Audio Engineering Society standardization effort in metadata started independently, and also adopted the approach of using Dublin Core. It was very encouraging to discover that the AES and the EBU had a common approach, and work is now in hand to ensure that the final AES document is as close to the EBU document Tech 3293 as possible.

Expression of the metadata

EBU Tech 3293 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. Dublin Core itself has been widely implemented in HTML and XML, and there is guidance documentation available from DCMI [2] on such implementations.

Conclusions

P/FRA met five times over a period of 18 months, visiting IRT, NAA and the BBC as well as meeting adjacent to IBC and AES meetings. During these meetings, archivists and engineers were both represented, in approximately equal numbers. As well as working on the standards documents, in each case we had technical tours at

Abbreviations

AES	Audio Engineering Society	NISO	National Information Standards Organization (USA)
CEN	<i>Comité Européen de Normalisation</i>	OCLA	Online Computer Library Center (USA)
DCMES	Dublin Core Metadata Element Set	SAM	Scandinavian Audiovisual Metadata group
DCMI	Dublin Core Metadata Initiative	SMPTE	Society of Motion Picture and Television Engineers (USA)
ISO	International Organization for Standardization	XML	Extensible Markup Language
HTML	HyperText Markup Language		



Richard Wright was educated at the University of Michigan (USA) and Southampton University (UK). Over the course of these studies, he obtained a BSc in Engineering Science (1967), an MA in Computer Science (1972) and a Ph D in Digital Signal Processing – Speech Synthesis (1988).

Dr Wright has worked in acoustics, speech and signal processing for US and UK Government research laboratories (1968-76), at the University College of London (1976-80; Research Fellow) and at the Royal National Institute for the Deaf (1980-90; Senior Scientist). He was the Chief Designer at Cirrus Research from 1990 to 1994 (acoustical and audiometric instrumentation).

Richard Wright has been the Technology Manager of BBC Archives since 1994. He is also the Head of EBU working group, P/FRA – Future Radio Archives, and of the EC-sponsored project PRESTO (Preservation Technology).

Marit Grimstad was educated at the Norwegian School of Librarianship, then spent a year studying Information Technology at the Norwegian School of Management, BI. This was followed by a one-year course in management for librarians.

Since the 1970s, Ms Grimstad has worked in the Radio Archive of the Norwegian Broadcasting Corporation, NRK. She became Head of the Radio Archive in 1989. Since 2000, she has been the project manager for Digital Radio Archives in NRK.

Marit Grimstad is a member of EBU working group, P/FRA; on Future Radio Archives. She is Head of NRK's metadata group and Head of SAM (Scandinavian Audiovisual Metadata group).



the host institution, and also shared our progress in radio archive digitization. These digitization projects are relevant, because as our archives become electronic files in a sea of servers or data tapes, **ONLY** the metadata will allow programme retrieval. Similarly, for electronic exchange, it is the metadata that will “make it all come right” rather than sowing the seeds of confusion as we move away from physical programme carriers and into the mass-storage age.

For the authors, it was pleasurable and very satisfying to benefit from the collective experience brought to the P/FRA table. One of the final recommendations of P/FRA was for the EBU to consider ways of continuing the exchange, pan-EBU, of information on radio digitization progress.

Bibliography

- [1] EBU Tech 3293: **EBU Core Metadata Set for Radio Archives**
http://www.ebu.ch/tech_t3293.html
- [2] **DCMI Dublin Core Metadata Initiative**
<http://dublincore.org/>
- [3] **EBU P/META Metadata Scheme**: to be published
- [4] **SMPTE Metadata Dictionary** as specified in SMPTE RP210a
- [5] **EBU/SMPTE Task Force final report**:
http://www.ebu.ch/pmc_es_tf.html#rep
- [6] SMPTE website
<http://www.smpte-ra.org/mdd/index.html>
- [7] SMPTE Standard 336M-2001 for Television : **Data Encoding Protocol using Key-Length-Value.**


Appendix A:

Example of metadata for a document: a web page

An electronic (web) document < <http://www.nla.gov.au/meta/> > is shown below:

home
http://www.nla.gov.au/meta/

Meta Matters



This Website is intended to help Web content providers improve the effectiveness of searching for information resources on the World Wide Web.

Difficulties in finding scattered Web resources have prompted the development of simplified metadata standards which could be used by authors, or Web content creators/publishers, to facilitate easier access for Web users.


The National Library of Australia encourages any initiative which seeks to implement standard metadata schemas such as the Dublin Core.

Introduction


Resources:

- [Discussion Lists](#)
- [FAQ](#)
- [Guides](#)
- [Creators](#)
- [Tools](#)
- [Links](#)
- [Reading list](#)
- [Australian Subject Gateways](#)

[Sitemap](#)
[Feedback](#)



[View metadata for
Meta Matters](#)



Copyright National Library of Australia
[Disclaimer](#)

© Commonwealth of Australia, 2000

A partial list of the metadata associated with this web document is given in the following *Table*:

Partial metadata for the web document shown on the previous page

Metadata Element	Scheme	Language	Content
DC.Identifier		en	http://www.nla.gov.au/meta/
DC.Creator		en	National Library of Australia
DC.Publisher		en	National Library of Australia
DC.Title		en	MetaMatters
DC.Description		en	This Website is intended to help Web content providers improve the effectiveness of searching for information resources on the Worldwide Web, by describing the metadata schemas available for use in Australia and their Australian implementations.
DC.Date	ISO8601	en	1999-05-24
DC.Type		en	Document

And shown below is the metadata that has been inserted in the <HEAD> HTML tag of the web page:

```

<!DOCTYPE HTML PUBLIC "-//W3C//DTD html 3.2 Final//EN">
<HTML>
  <HEAD>
    <TITLE>Meta Matters | Metadata for Meta Matters</TITLE>
    <META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=iso-8859-1">
    <META NAME="DC.Identifier" LANG="en"
    CONTENT="http://www.nla.gov.au/meta/lists.html">
    <META NAME="DC.Creator " LANG="en" CONTENT="National Library of
    Australia">
    <META NAME="DC.Creator.Email " LANG="en" CONTENT="metadata@nla.gov.au">
    <META NAME="DC.Publisher " LANG="en"
    CONTENT="National Library of Australia">
    <META NAME="DC.Title" LANG="en" CONTENT="MetaMatters Discussion Lists">
    <LINK REL="schema.DC" HREF="http://mirror.nla.gov.au/dc/elements/1.0/">
    <META NAME="DC.Subject" LANG="en" CONTENT="metadata creation">
    <META NAME="DC.Subject" LANG="en" CONTENT="metadata architectures">
    <META NAME="DC.Subject" LANG="en" CONTENT="resource discovery">
    <META NAME="DC.Subject" LANG="en" CONTENT="subject gateways">
    <META NAME="DC.Description" LANG="en"
    CONTENT=" The MetaMatters site provides links metadata discussion lists,
    to which membership is open to any individual interested in the use of
    metadata in Australia. ">
    <META NAME="DC.Language" SCHEME="RFC1766" LANG="en" CONTENT="en">
    <META NAME="DC.Coverage" LANG="en" CONTENT="Commonwealth">
    <META NAME="AGLS.Function" LANG="en"
    CONTENT="Recordkeeping Standards - Advice (NAA Functions Thesaurus)">
    <META NAME="DC.Date" SCHEME="ISO8601" LANG="en" CONTENT="1999-05-24">
    <META NAME="DC.Type" LANG="en" CONTENT="Document">
    <META NAME="DC.Format" SCHEME="IMT" LANG="en" CONTENT="text/html">
  </HEAD>

```