Metadata in a distributed information environment: Interoperability as recombinant potential

Lorcan Dempsey

OCLC/SCURL pre-IFLA conference, 15/16 Aug 02
Overview

• What is metadata?
• Some application examples
• Recombinant potential
Summary

- If it moves it will be covered in metadata.
- The value of future digitization and distributed information initiatives will crucially depend on the creation and management of a variety of types of metadata.
- Metadata will allow machines and users behave smarter.
What is metadata?
Your 634,000 search results for "metadata"

Dublin Core Metadata Initiative (DCMI)
... Graphic for the DC-2002 International Conference. The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online ...
dublincore.org/

Metadata at W3C
Metadata and Resource Description. Metadata is machine understandable information for the web. The W3C Metadata Activity addressed ...
www.w3.org/Metadata/

Digital Libraries: Metadata Resources
... GENERAL RESOURCES AND INDICES. Metadata is data about data. The term refers to any data used to aid the identification, description ...
www.ifl.org/ll/metadata.htm

UKOLN Metadata
Metadata. UKOLN. Projects. We are participants in the following projects which relate to resource description: agentcities ... What is metadata? Metadata ...
www.ukoln.ac.uk/metadata/
“The metadata necessary for successful management and use of digital objects is both more extensive than and different from the metadata used for managing collections of printed works and other physical materials.

While a library may record descriptive metadata regarding a book in its collection, the book will not dissolve into a series of unconnected pages if the library fails to record structural metadata regarding the book's organization, nor will scholars be unable to evaluate the book's worth if the library fails to note that the book was produced using a Ryobi offset press.
The same cannot be said for a digital version of the same book. Without structural metadata, the page image or text files comprising the digital work are of little use, and without technical metadata regarding the digitization process, scholars may be unsure of how accurate a reflection of the original the digital version provides.

For internal management purposes, a library must have access to appropriate technical metadata in order to periodically refresh and migrate the data, ensuring the durability of valuable resources.”
Beyond discovery retrieval, use, management
THE MANAGEMENT AND AUTOMATIC PRODUCTION OF METADATA

J. O. Drewery and J. L. Riley

BBC R&D, UK

ABSTRACT

Metadata, if standardised, provides a means to help people work together to produce programmes for broadcast or internet delivery. This paper reviews the benefits of adopting metadata and identifies the particular kinds that are needed at different stages of the production chain. It then discusses the management of metadata, some opportunities for generating it automatically and the possibility of using agent technology to aid this process.

INTRODUCTION

The advent of satellite, cable and internet delivery, as well as the migration to digital engineering, has caused broadcasters to consider more efficient means of creating programmes for an increased number of outlets. The introduction and management of standardised information about programmes – referred to as metadata [1, 2] – is expected to play a key role in this aim. Provided that the bulk of metadata can be generated automatically and user interfaces are easy to use, we can also expect that a greater proportion of version control; and the need for unique identifiers for programmes and source material so that they can be referred to unambiguously.

During the past year, there has been considerable effort to ensure that the BBC work is not regarded in isolation. The EBU's P/Meta group has begun the task of examining other broadcasters' needs and developing a broader version of SMEF. The 'top 60' element descriptions of broadcaster-related metadata have been added to the SMPTE video metadata dictionary [3]. A primary aim is to seek international standardisation as a plank for
List of Records

- Click on a title to see the detailed record.
- Click on a checkbox to mark a record to be e-mailed or printed in Marked Records.
- Refine your search results: 1575  (Show me how)

1. Managing semantic content for the Web
   Author: Sheth, A.; Bertram, C.; Avant, D., and others
   Source: IEEE Internet Computing (IEEE)6, no. 4, (July-Aug. 2002) : 80-7
   Doc. Type: Journal Paper Libraries: 381

2. Managing scientific metadata using XML
   Author: Ruixin Yang; Kafatos, M.; Wang, X.S.
   Source: IEEE Internet Computing (IEEE)6, no. 4, (July-Aug. 2002) : 52-9
   Doc. Type: Journal Paper Libraries: 381

3. NewsML Toolkit
   Author: Compton, D.
   Source: Interchange (Int. SGML/XML Users' Group)8, no. 2, (June 2002) : 14-17
   Doc. Type: Journal Paper Libraries: 407

4. An experimental study on content-based image classification for satellite image databases
   Author: Holowczak, R.D.; Artigas, F.J.; Soon Ae Chun, and others
   Doc. Type: Journal Paper Libraries: 407

5. What is the Open Archives Initiative?
   Author: Morgan, E.L.
   Source: Interchange (Int. SGML/XML Users' Group)8, no. 2, (June 2002) : 18-22
   Doc. Type: Journal Paper Libraries: 407

6. Using RSS for syndication
   Author: Dodds, L.
   Source: Interchange (Int. SGML/XML Users' Group)8, no. 2, (June 2002) : 11-14
   Doc. Type: Journal Paper Libraries: 407

7. Building a bioinformatics ontology using OIL
   Author: Stevens, R.; Goble, C.; Horrocks, I., and others
   Doc. Type: Journal Paper Libraries: 291

8. OILING the way to machine understandable bioinformatics resources
   Author: Stevens, R.; Goble, C.; Horrocks, I., and others
   Source: IEEE Transactions on Information Technology in Biomedicine (IEEE)6, no. 2, (June 2002) : 129-34
   Doc. Type: Journal Paper Libraries: 291

9. The languages of the semantic Web
   Author: Ogbuji, U.
   Source: New Archit (CMP Media Inc)7, no. 6, (June 2002) : 30-3
   Doc. Type: Journal Paper Libraries: 93
Metadata is data in the system which....

... allows people and machines do smarter things.
Metadata is data which supports operations on resources ....

- Know what is available
- Know what can be done with something
  - Technically
  - Policy
- Know how to look after something over time
- Know how to pass things between processes
- Know how to package/disaggregate ..

... and relieves the user of having to have advance knowledge of the characteristics or existance of the resource.
Some application examples

- The portal
- Digital content management
- E-prints UK
- Recombinant library services
Some entities

- users
- institutions
- rights
- collections
- services
- information objects
- schemas
- policies
- Transactions/events
- terms
Library portal

- How the library mediates the engagement of users and resources in a network environment
When you search in DAIster, you're searching a wide variety of collections from a wide variety of institutions. These institutions have made the records of their digital resources available to us, and we have gathered and aggregated them into the DAIster service.

The following is a list of these collections, including a description of each one, and indicating how many records we have gathered from each collection. (The number of records may be fewer than the number the institution makes available -- we are making only those that point to a digital resource available.)

Currently, we are serving 329919 records from 89 institutions.

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

Academia Sinica (2 records)
http://www.isis.sinica.edu.tw/

The Institute of Information Science at Academia Sinica in Taiwan provides access to archives of texts designed to analyze modern Chinese and Formosan.

Ackerman Archives (147 records)
http://www.hrav.com/aa/

The Ackerman Archives contains digital library quality information in industry standard form, namely the Harvey Ladev Williams II papers, the Harvey Ladev Williams II World War I letters and a 1918 Italian Front photograph album. There are also family stories and images that have been OA enabled.

Advanced Knowledge Technologies (AKT) EPrints Archive (20 records)
http://eprints.aktors.org/

Members of the Advanced Knowledge Technologies (AKT) project have submitted papers to this eprints archive. The AKT project aims to develop and extend a range of technologies providing integrated methods and services for the capture, modelling, publishing, reuse and management of knowledge. This archive is running on eprints.org open archive software, a freely distributable archive system available from eprints.org.

Alex Catalogue of Electronic Texts (15 records)
http://www.infomotions.com/alex/

The Alex Catalogue of Electronic Texts is a collection of public domain documents from American and English literature as well as Western philosophy. The catalogue's primary purpose is to assist in demonstrating the concept underlying a process of understanding using methods from natural language. The catalogue's intended audience is
Although these technological components are referred to in the ensuing discussion, the scope of the paper does not include a detailed description of each. For more information about the technology, please see www.exlibris.co.il/MetaLibView.html.

Application Level:

Technology Level:

![Diagram of MetaLib](http://www.caslin.cz.cz:7777/caslin01/sbornik/k/metalib.html)

Figure 1: MetaLib, a Web application that is based on technological components and relies on a knowledge base.

Several issues are involved in the goals that MetaLib aims to achieve:

- **Connectivity**: How will the software connect to heterogeneous resources?
- **Queries**: How can the software achieve a single query interface?
- **Data manipulation**: How will the data received from external resources be processed?
- **The "fork in the road"**: What options will be available to the user once an item is found as the result of a search?
- **The knowledge base**: What knowledge does the software require to accomplish all the tasks listed above?
OCLC

- In planning
  - Portal utilities
    - ILL policy directory
    - Collection and service description (feasibility?)
  - Rights and resolution service

- In research
  - Terminology services
Digital content management

• A life cycle approach
  – Minimize long-term costs
  – Maximize value of investment through use

• Metadata crucial at all phases

• How ‘collectible’/’ingestible’ is an object

• Preservation/METS/OAIS

DUBLIN, Ohio, June 11, 2002—The Working Group on Preservation Metadata, an initiative jointly sponsored by OCLC and RLG, has released a Metadata Framework to Support the Preservation of Digital Objects (PDF 1155K), a new report available on the OCLC web site.

The report is a comprehensive guide to preservation metadata that is applicable to a broad range of digital preservation activities. Preservation metadata is the information infrastructure necessary to support processes associated with the long-term retention of digital resources, and is an essential component of most digital preservation systems.

The report represents the consensus of leading experts and practitioners comprising the working group, and is intended for use by organizations and institutions managing, or planning to manage, the long-term retention of digital resources.

The working group based its work on preservation metadata element sets developed by several leading institutions and organizations in the digital preservation community, as well as the Open Archival Information System (OAIS) reference model. OAIS is a conceptual framework that articulates the primary functional components of a digital archive and establishes concepts and terminology for describing and comparing architectures and data models.

The report follows on the working group’s earlier white paper Digital Objects: A Review of the State of the Art (PDF 318K), which defined and discussed the concept of preservation metadata, reviewed current thinking and practice in the use of preservation metadata, and identified starting points for consensus-building activity in this area.

*The working group’s report illustrates the value of cooperatively resolving the challenges of digital
The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation.

News & Announcements

- METS Editorial Board Formed (May 29, 2002)
- New Version of METS Schema released (February 8, 2002)

METS Schema, & Documentation

- METS Overview & Tutorial
- METS Schema 1.0 (zeta)
- METS Documentation [Created with XML Console software]
- Xlink Schema

METS Schema (Proposed New Version)

- METS Schema 1.1
  
  Note: this is NOT the official current version. The official current version is METS Schema 1.0 (zeta) listed above.

METS Example Documents
Mets

- **Descriptive**: facilitating resource discovery and identification
- **Administrative**: supporting resource management within a collection
- **Structural**: binding together the components of complex information objects
E-prints UK project

- Part of JISC’s FAIR programme
- Disclosure of institutional assets
  - Institutional repositories
- Partners
  - UKOLN/RDN
  - University of Southampton
  - OCLC
- Make metadata do more work!
- Add value to metadata in a network of distributed services
E-prints UK

Institutional e-print archives

Non-institutional e-print archives

Personal e-print archives

SOAP

OAI-PMH

E-Prints UK

Subject classification service

Name authority service

Citation analysis service

SOAP

Javascript HTTP

RDN Gateway
**Metadata switch**

- **Explore** a set of services which add value to metadata by leveraging OCLC expertise and cooperative processing model.

- **Services in an open environment of use**
  - Harvesting
    - Aggregation/fusion
  - Name services
  - Knowledge organization services
  - Schema transformation services
Aggregation and fusion
Schema transformation
Knowledge organization
Names
Interoperability as recombinant potential

- Disaggregating scholarly publishing
  - Linking, Identifiers
- ‘Play’ learning objects
  - Packaged
- Federated searching
  - Fusing metadata
- Processing content
- Ingesting content
- ‘Plugging in’ services

Examples
- Can I add a document to a repository?
- Can I add a repository to a distributed query?
- Can I fuse metadata from one repository with another?
- Can I aggregate these resources into a learning package?
In conclusion

- Way beyond discovery
- Realise value in distributed services and digitization
- New potential for shared and third party services
- Correct approach integral to effective conduct of learning and research
- Make metadata work!