Metadata Overview

**Public Broadcasting
Metadata Model Project**
Working Group Meeting
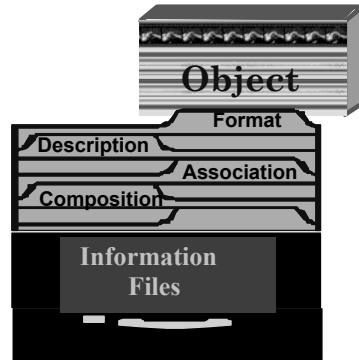April 24, 2002
Arlington, VA
**Thom Shepard, MLS, Applied Technology
WGBH Educational Foundation**

Outline of Presentation

- **Selected Glossary Terms**
- **Metadata for Broadcast Media**
- **Metadata Standards**
- **WGBH Clean-up Initiative**

What is Metadata?

- **Ontology**
- **Property sets**
- **Resource discovery handles**

What is a Resource?

- **Examples:**
  - **Electronic document**
  - **Image**
  - **Service**
  - **Collection of other *resources***
  - ***Even people!***
- **Anything with an identity**

What's an Identifier?

- **a key**
- **unambiguous reference**
- **a metadata kernel**

Name some Identifier Standards

- **Digital Object Identifier** (DOI)
  - **property rights in a digital environment**
- **Unique Material Identifier** (UMID)
  - **A/V content units in a clip or shot**
- **Universal Program Identifier** (UPID)
  - **IDs a single piece of completed content**
- **International Standard Audiovisual Number** (ISAN)
  - **numbering system to identify audiovisual "works"**

ISAN
THEREFORE
I AM

What is Bibliographic Description?

- **The data elements that identify a particular work as unique.**
  - ID
  - Title
  - Ownership
  - Edition
  - Material
  - Publication information
  - Description
  - Notes

What is Data Modeling?

- Determining data requirements
- Documenting relationships
- ER diagram
- Phases
  - Logical
  - Physical

Where's an Example of a Data Model?

- **SMEF-DM (BBC)**
  - **Data model for Media Asset Management**
  - **Life Cycle**
  - **Available for free**
  - **Includes follow-up questionnaire**

  http://www.bbc.co.uk/guidelines/smef/

What is a Schema?

- **Rules for encoding information**
- **Supports communities of users**
- **Standard way of defining elements & relationships**

What is Content?

- **Essence**
- **Physical object**
  - **Shelf space**
  - **Shelf life**
- **Virtual object**
  - **File space**
  - **Filename**
  - **Directory location**

What is Content?

- **Essence**
- **Physical object**
  - Shelf space
  - Shelf life
- **Virtual object**
  - File space
  - Filename
  - Directory location

```
A:\DIR

  Directory of A:\

FR. NAME EXT.      PRIZE    CR-DATE      TIME

PERSROST RPT       14221   7-18-96      8:34a
SUPAUDIT DOC         983   7-19-96     11:03a
SUPUPDAT EXO       83362   7-21-96      3:05p
KIV-PRT  MON         684   7-22-96      7:27p
FOODSERV UPD       25642   7-24-96      9:52a
DIN-MENU DLY        6363   7-25-96      8:36a
DISROFFR RPT         737   7-25-96     10:23a
IM COORD DOC       23323   7-25-96     11:15p
SUPUDAT  EX2        1483   7-26-96      5:31a
EVALS-DU DP2        8864   7-26-96      4:05p
PERS-UPD NEW        9958   7-27-96      7:45a
PAYROLL  LST       31487   7-28-96      3:00p
SUP-INVT EX2        4343   7-28-96      6:24p
IMMS-RPT LTR        6813   8-02-96      9:47a
DIV-NOTE F         10621   8-03-96      1:23a
IM-AWRPT MON       17453   8-03-96      4:29a
DIV-CMT  WK4        1690   8-06-96     10:18a
FIREBILL XD         5090   8-06-96      1:00p
WORK     XD          821   8-16-96      9:18a
ETC.
ETC.
FILE 50

       58 FILES(S)    161435 bytes free
```

What's Rights Management?

- **Technologies**
- **Tools**
- **Procedures**
- **Protecting intellectual property**

What are some Rights Standards?

- **DRM Framework**
  - Functions & Information Architectures
  - expressed as ODRL
- **Open Digital Rights Language**
  - Vocabulary for terms & conditions
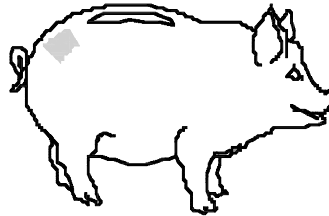- **XrML**
  - Rights language as XML / RDF namespace

What is an Asset?

- **Essence**
- **Metadata**
- **Rights**

What is a Repository?

- **Collection of assets**
- **Databases with a common search engine**

What is Harvesting?

- **Gathering**
- **Indexing**
- **Accessing**

- **Protocol**
  - **Rules for compatibility**

Metadata for Broadcast Media

- **Hierarchical & Relational**
- **Modular**
- **Time-based**
- **Migrations, Ambiguities & Transformations**
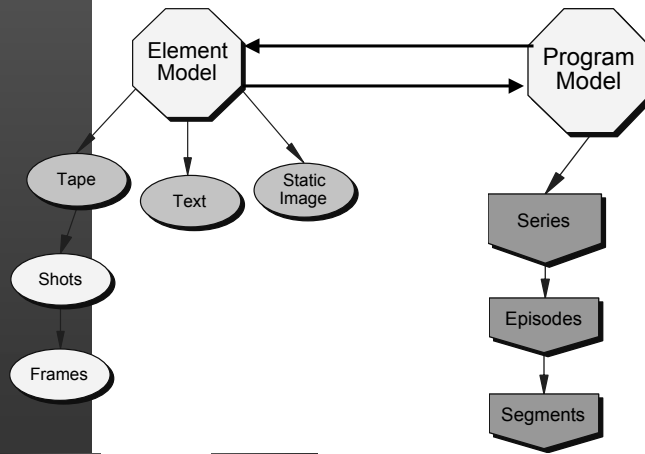- **Hybrid**
- **Exchange Standards**

Relational versus Hierarchical

- **Relational**
  - **Tables, Columns, Rows**
  - **Key fields**
  - **Metadata by association**
  - **Example: Program to its Elements**
- **Hierarchical**
  - **Nested property sets**
  - **Inheritance**

Physical & Intellectual Control as Hierarchy

Physical Control   Intellectual Control

Element
Model ⟷ Program
Model

Tape   Text   Static
Image

Shots

Frames
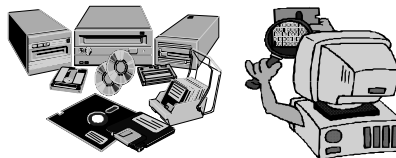
Series

Episodes

Segments

Metadata for Broadcast Media

- **Hierarchical & Relational**
- **Modular**
- **Time-based**
- **Migrations, Ambiguities & Transformations**
- **Hybrid**
- **Exchange Standards**

WGBH Divisions

- **Television**
- **Radio**
- **Interactive (Web)**
- **Information Technology**

- **Administration**
- **Legal**
- **Enterprise**
- **Design**
- **Archives**
- **Media Library**

Metadata for Broadcast Media

QuickTime™ and a
GIF decompressor
are needed to see this picture.

- **Hierarchical & Relational**
- **Modular**
- **Time-based**
- **Migrations, Ambiguities & Transformations**
- **Hybrid**
- **Exchange Standards**

Time-Based Media: Characteristics

- **Data flow**
- **Duration**
- **Temporal composition**
- **Synchronization**

Time-Based Media: Metadata

- **Segmentation**
  - **Small units**
  - **Shots**
  - **Description**
  - **Subjective**
- **Stratification**
  - **"Chunks"**
  - **Frames**
  - **Keywords**
  - **Objective**

Metadata for Broadcast Media

- **Hierarchical & Relational**
- **Modular**
- **Time-based**
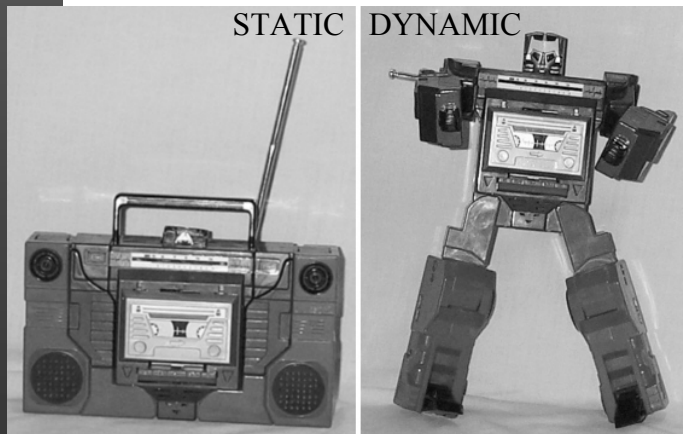- **Migrations, Ambiguities & Transformations**
- **Hybrid**
- **Exchange Standards**

Migrations & Ambiguities

- **Preservation Metadata**
- **Proxies as Assets**
- **Items as Collections**
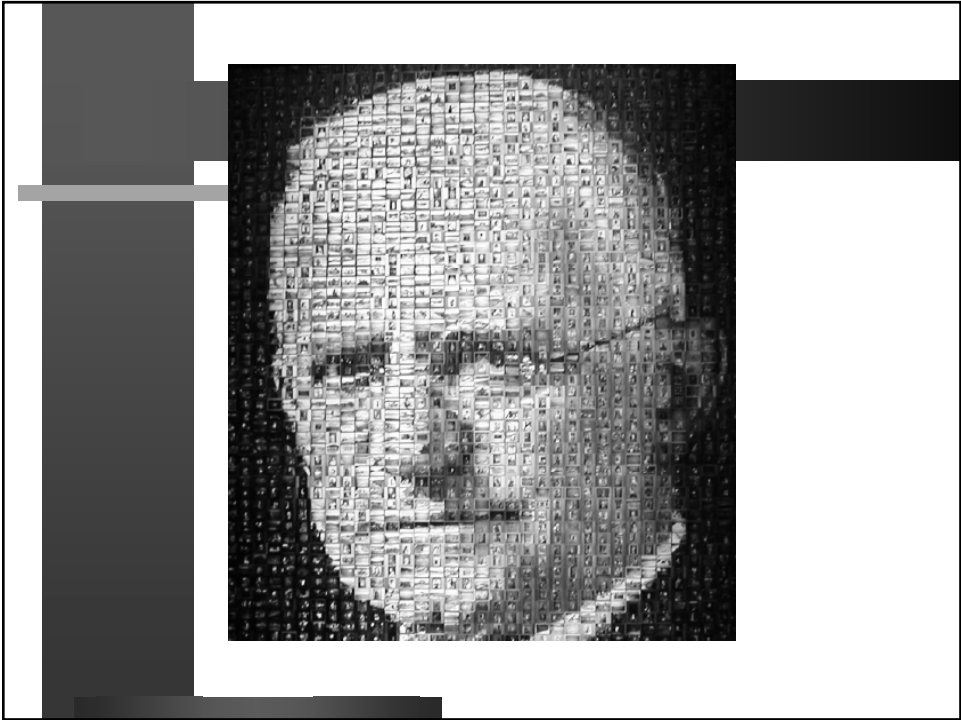
Transformations

STATIC   DYNAMIC

Metadata for Broadcast Media

QuickTime™ and a
Photo - JPEG decompressor
are needed to see this picture.

- Hierarchical & Relational
- Modular
- Time-based
- Migrations, Ambiguities & Transformations
- Hybrid
- Exchange Standards

Hybrid: Materials
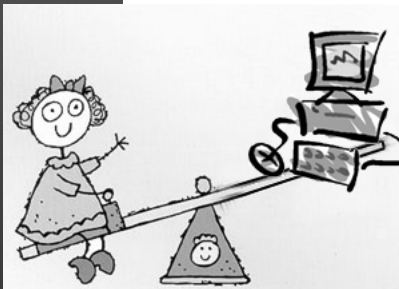
- **Physical & Virtual**
- **Analog & Digital**
- **Mixed media**

- **Embedded & Standalone Metadata**

Metadata for Broadcast Media

- **Hierarchical & Relational**
- **Modular**
- **Time-based**
- **Migrations, Ambiguities & Transformations**
- **Hybrid**
- **Exchange Standards**

Hybrid: Roles

- **Who describes it?**
- **What extracts it?**
- **How is it collected?**
- **When, how and by whom is it normalized?**

Metadata for Broadcast Media

- **Hierarchical & Relational**
- **Modular**
- **Time-based**
- **Migrations, Ambiguities & Transformations**
- **Hybrid**
- **Standards**

Exchange Standards: XML & RDF

- **eXtensible Markup Language**
  - **Nesting of values, attributes**
  - **Self-describing**
  - **Types**
- **RDF**
  - **Resource Description Framework**
  - **Modeling language within XML**
  - **Namespaces**
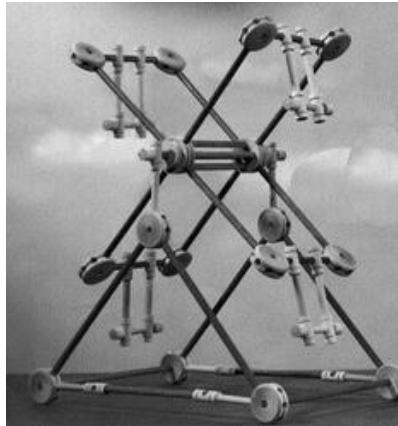
Asset Expressed as XML

```
<ASSET>
   <METADATA>
       <Title>Metadata Glossary & Overview</Title>
       <Type>Presentation</Type>
       <Format>Powerpoint</Format>
       <Description>A bold attempt to explain everything
   about metadata in 30.</Description>
       <Rights>All rights in perpetuity</Rights>
   </METADATA>
   <CONTENT>CPB Metadata Model.ppt</CONTENT>
</ASSET>
```

Metadata Expressed as XML / RDF

```xml
<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about="http://doc">
     <author> Joe Smith </author>
  </rdf:Description>
</rdf:RDF>
```

Metadata Expressed as RDF Namespaces

```xml
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dc="http://purl.org/dc/elements/1.0/">
  <rdf:Description rdf:about="http://doc">
   <dc:creator> Joe Smith </dc:creator>
  </rdf:Description>
</rdf:RDF>
```

What is the Dublin Core?

- **<u>Simple</u> element set for describing a wide range of networked resources**
- **15 metadata elements**
- **International and cross-disciplinary**
- **Stable since 1996**
- **NISO Standard: Z39.85-200**

Grouping DC Elements

<u>**Content**</u>
- **Coverage**
- **Description**
- **Type**
- **Relation**
- **Source**
- **Subject**
- **Title**

<u>**Intellectual Property**</u>
- **Contributor**
- **Creator**
- **Publisher**
- **Rights**

<u>**Versioning**</u>
- **Date**
- **Format**
- **Identifier**
- **Language**

## Characteristics of Dublin Core

- **Easy to understand**
- **Commonly understood semantics**
- **International**
- **Extensible**

Communities of Dublin Core

- **Libraries**
- **Archives**
- **Museums**
- **Publishing**
- **Industry**
- **Government**
- **Dare we add, Broadcasting?**

Dublin Core Relevance to Broadcasting

- **Definitions**
- **Meta tags for resource discovery**
- **Database modeling**
- **Qualifier attributes**
- **XML**
- **Harnessing free-form data**
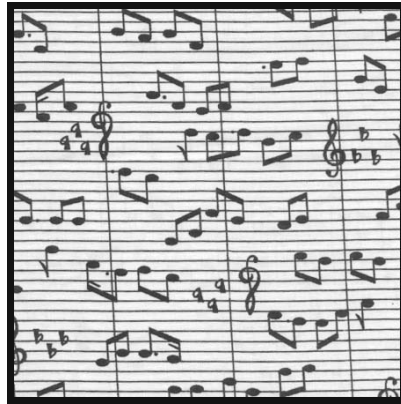
Simple versus Qualified DC

- **User-friendly**
- **General-purpose**
- **Insufficient for Program Metadata**
  - **Titles**
  - **People & roles**
  - **Dates**
  - **Etc**
- **XML Schema**

Simple Dublin Core


```
<dc xmlns="http://purl.org/dc/elements/1.1/"
```

Simple versus Qualified DC

- **Qualified:**
  - **increased specificity**
  - **referencing existing schemes or controlled vocabularies**
  - **DC community registers qualifiers**
- **Standard Qualifiers**
  - **Element Refinement**
  - **Encoding Scheme**
- **Proprietary Qualifiers**
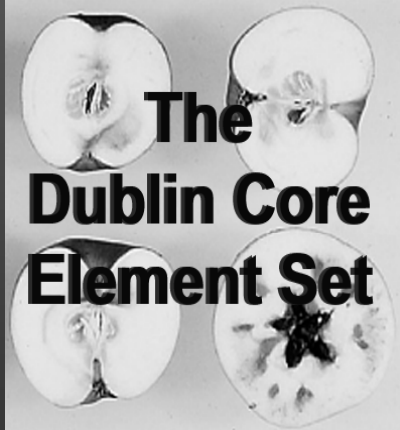 - **Community DTD**
  - **trade-off: may hinder interoperability**

Dublin Core as a Language

QuickTime™ and a
GIF decompressor
are needed to see this picture.

Standards

The Dublin Core Element Set

- METS
- SCORM
- LOM / IMS
- OAI
- MPEG-7
- MARC

Standards

QuickTime™ and a
Photo - JPEG decompressor
are needed to see this picture.

- METS
- SCORM
- LOM / IMS
- OAI
- MPEG-7
- MARC

METS

- **Metadata Encoding & Transmission Standard**
- **XML schema**
- **Library of Congress**
- **Digital Library Federation**
- **Digital Library Objects within a Digital Library**
- **Maps to Dublin Core**

- **Categories of Metadata**
  - **Descriptive Metadata**
  - **Administrative Metadata**
  - **File Groups**
  - **Structural Map**

SCORM

- **Sharable Content Object Reference Model**
- **eLearning**
- **Department of Defense**

- **Dynamic learning**
  - **Custom-assembled**
  - **Personal pace**
  - **Personal needs**
- **Content packaging**

IMS / LOM

- **IEEE**
- **Instructional Management Systems**
- **Learning Technology Standards Committee**
- **Learning Object Metadata Working Group**

http://www.imsproject.org/

Learning Object

"Any entity, digital or non-digital, that may be used for learning, education, or training."

IMS / LOM to DC: Examples

Title
   general.title
Subject
   general.keywords
Description
      general.description
Contributor
      lifecycle.contribute
   lifecycle.contribute.role
Format
   technical.format

Date
      lifecycle.contribute.date
Source
   relation.resource
   relation.kind ="IsBasedOn
Language
      general.language
Coverage
      general.coverage
Rights
   rights.description

OAI (Open Archives Initiative)

- OAI Protocol for "Metadata Harvesting"
- XML
- Resource discovery over distributed collections

- Mandates Simple DC
- Supports other schemes
- May inspire creation of community-specific XML schemes

MPEG-7

- Structured descriptions of multimedia content
- XML
- Supports granularity
- Open standard
- Enables binary searching over the web
  - Notes can find a song
  - Sketch can find an image
  - Six degrees of Kevin Bacon

Dublin Core "Paths" in MPEG-7

CreationMetaInformation.Creation.Title.TitleText
CreationMetaInformation.Creation.Creator (role="creator")
CreationMetaInformation.Creation.CreationDescription
CreationMetaInformation.Creation.Creator (role="publisher")
CreationMetaInformation.Creation.Creator (role="contributor")
CreationMetaInformation.Creation.CreationDate
CreationMetaInformation.Classification.Genre
MediaInformation.MediaProfile.MediaFormat.FileFormat
MediaInformation.MediaIdentification.Identifier
CreationMetaInformation.Classifcation.Language.LanguageCode
UsageMetaInformation.Rights.RightsID

MPEG-7 "Harmony Project"

- **Multimedia Content Description Interface**
- **Dublin Core**
- **MPEG-7**
- **XML / RDF Namespaces**

MARC (ANIM & AACR2)

- **Archival Moving Image Materials: a Cataloging Manual**
- **Anglo-American Cataloguing Rules**
- **Dublin Core Crosswalk**
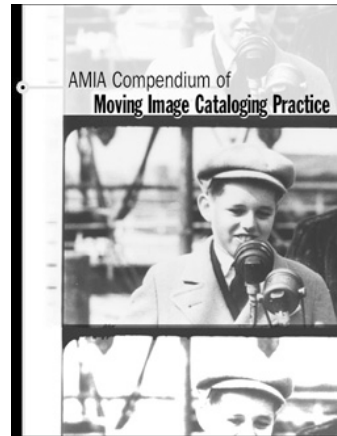
"Things should be as simple as possible, but not simpler."

Albert Einstein

More information about Dublin Core

- **Dublin Core Home Page**
  - http://purl.oclc.org/dc
- **Jane Hunter Homepage**
  - http://archive.dstc.edu.au/RDU/staff/jane-hunter.html
- **Grace Agnew, Dan Kniesner**
  - **"Dublin Core Application Profile for Digital Video"**
  - http://www.vide.net/workgroups/videoaccess/resources/vide_dc_userguide_20010909.pdf
- **Consortium for the Computer Interchange of Museum Information**
  - **"CIMI Guide to Best Practices - Dublin Core 1.1"**
  - http://www.cimi.org/old_site/documents/meta_bestprac_v1_1_210400.pdf