Using Multimedia Metadata

- Semantic Gap
- Manual textual annotation
  - Subjective process
- Meta data generation
- MP3
- EXIF
- Dublin Core
- MPEG-7
  - MPEG-7 and the semantic gap
  - MPEG-7 Annotation
  - MPEG future, MPEG-21
- MPEG-7 and indexing
Metadata is literally „data about data“
- Often a text description of an image or audio clip can be provided
- Metadata deals with the content structure and similarities of the data
- Metadata generation is an important aspect of multimedia database, because directly queering the data can be extremely difficult

What is the Problem? Semantic Gap
- Today people cannot easily create, find, edit, share, and reuse media
- Computers don’t understand media content
  - Media is opaque and data rich
  - We lack structured representations
- Without content representation (metadata), manipulating digital media will remain like word-processing with bitmaps
Meta Data

- Meta Data represented as text, key words

- Text based retrieval
  - Works by adding brief description combined with some structured database
  - Some images are very difficult to describe in text, subjective

- Manual textual annotation
  - Images make no attempt to tell us what they are about
  - It is very difficult to express in words what a work of art is about
  - Getty Information Institute, which has contributed a great development of metadata standards for art collections, average time seven minutes per image
  - Manual annotation is very time consuming
The process of describing an image is highly subjective
An image can mean different things to different people
It can mean different things to the same person at different times
Mismatch between the user's terms and those of the indexer

Since individual objects will have their associative metadata
- Need to be maintained along with the original data
- Metadata needs to be updated when the raw data is changed
- When metadata exist, we need to know how to store and update it
Meta Data

- Manual generations
- Semi-automatic
- Automatically - *(Signature?)*
  - texture for images
  - frequencies for audio
  - font size for texture
  - motion direction and lighting for videos
  - speech-speaker place time
  - video clip, camera motion and lighting

Standards Overview

- Why do we need multimedia standards?
  - Reliability
  - Scalability
  - Interoperability
  - Layered architecture
MP3 metadata

- ID3 is a metadata container most often used in conjunction with the MP3 audio file format.
- It allows information such as the title, artist, album, track number, or other information about the file to be stored in the file itself.

<table>
<thead>
<tr>
<th>ID3 Tag v1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song title</td>
</tr>
<tr>
<td>Artist</td>
</tr>
<tr>
<td>Album</td>
</tr>
<tr>
<td>Year</td>
</tr>
<tr>
<td>Comment</td>
</tr>
<tr>
<td>Genre</td>
</tr>
</tbody>
</table>

Digital pictures and metadata

- When taking a picture, the digital equipment can automatically embed information like date and time.
- More recently, some digital cameras have GPS incorporated, and more sophisticated information is being embedded on the photos.
- Typically this metadata is directly embedded on the file.
- Both JPEG and TIFF file formats foresee the possibility of having extra information embedded.
EXIF

- EXIF (EXchangeable Image File format, www.exif.org/) takes advantage of metadata

- EXIF is a specification for using in image file format used by digital cameras

- EXIF stands for Exchangeable Image File Format, and is a standard for storing interchange information in image files, especially those using JPEG compression

EXIF

- Most digital cameras now use the EXIF format

- The format is part of the DCF standard created by JEITA http://www.jeita.or.jp/

- Encourage interoperability between imaging devices
### IFD 0

<table>
<thead>
<tr>
<th>Entry</th>
<th>Tag</th>
<th>Value</th>
<th>Meaning</th>
<th>Format</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make</td>
<td>0100</td>
<td>FUJIFILM</td>
<td>ascii string</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>0110</td>
<td>FinePix40i</td>
<td>ascii string</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td>0112</td>
<td>1</td>
<td>0 degrees</td>
<td>unsigned short</td>
<td>1</td>
</tr>
<tr>
<td>X Resolution</td>
<td>0114</td>
<td>72/1</td>
<td>72.00</td>
<td>unsigned rational</td>
<td>1</td>
</tr>
<tr>
<td>Y Resolution</td>
<td>0116</td>
<td>72/1</td>
<td>72.00</td>
<td>unsigned rational</td>
<td>1</td>
</tr>
<tr>
<td>Resolution Unit</td>
<td>0128</td>
<td>2</td>
<td>inch</td>
<td>unsigned short</td>
<td>1</td>
</tr>
<tr>
<td>Software</td>
<td>0131</td>
<td>Digital Camera FinePix40i Ver1.39</td>
<td>ascii string</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Date/Time</td>
<td>0132</td>
<td>2000:08:04 18:22:57</td>
<td>ascii string</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>YCbCr Positioning</td>
<td>0213</td>
<td>2</td>
<td></td>
<td>datum point</td>
<td>unsigned short</td>
</tr>
<tr>
<td>Copyright</td>
<td>8298</td>
<td></td>
<td>ascii string</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Exif Offset</td>
<td>8769</td>
<td>250</td>
<td></td>
<td>unsigned long</td>
<td>1</td>
</tr>
</tbody>
</table>

### EXIF Sub IFD

<table>
<thead>
<tr>
<th>Entry</th>
<th>Tag</th>
<th>Value</th>
<th>Meaning</th>
<th>Format</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>F Number</td>
<td>8220</td>
<td>281:100</td>
<td>2.80</td>
<td>unsigned rational</td>
<td>1</td>
</tr>
<tr>
<td>Exposure Program</td>
<td>8223</td>
<td>5</td>
<td>program portrait</td>
<td>unsigned short</td>
<td>1</td>
</tr>
<tr>
<td>ISO Speed</td>
<td>8273</td>
<td>200</td>
<td>ISO 200</td>
<td>unsigned short</td>
<td>1</td>
</tr>
<tr>
<td>EXIF Version</td>
<td>85D0</td>
<td>0.0.0.0.0.0.0.0</td>
<td>undefined</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Date/Time Stamp</td>
<td>8601</td>
<td>2000:08:04 18:22:57</td>
<td>ascii string</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Date/Time Digital</td>
<td>8604</td>
<td>2000:08:04 18:22:57</td>
<td>ascii string</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Component Configuration</td>
<td>9193</td>
<td>01:02:03:00</td>
<td></td>
<td>undefined</td>
<td>4</td>
</tr>
<tr>
<td>Compressed FinePrint</td>
<td>0102</td>
<td>1210</td>
<td>1210</td>
<td>unsigned rational</td>
<td>1</td>
</tr>
<tr>
<td>Master Speed Value</td>
<td>9213</td>
<td>5.50</td>
<td>5.50</td>
<td>signed rational</td>
<td>1</td>
</tr>
<tr>
<td>Avenue Value</td>
<td>9202</td>
<td>200:100</td>
<td>200:100</td>
<td>unsigned rational</td>
<td>1</td>
</tr>
<tr>
<td>Brightness Value</td>
<td>9205</td>
<td>200:100</td>
<td>200:100</td>
<td>unsigned rational</td>
<td>1</td>
</tr>
<tr>
<td>Exposure Bias Value</td>
<td>9206</td>
<td>0.00</td>
<td>0.00</td>
<td>signed rational</td>
<td>1</td>
</tr>
<tr>
<td>Max Aperture Value</td>
<td>9208</td>
<td>0.00</td>
<td>0.00</td>
<td>unsigned rational</td>
<td>1</td>
</tr>
</tbody>
</table>
### Maker Note (FujiFilm)

<table>
<thead>
<tr>
<th>Entry</th>
<th>Tag</th>
<th>Value</th>
<th>Meaning</th>
<th>Format</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>0000</td>
<td>30 31 33 30</td>
<td>undefined</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>1000</td>
<td>NORMAL</td>
<td>ascii string</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Sharpness</td>
<td>1001</td>
<td>3</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>White Balance</td>
<td>1002</td>
<td>0</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Flash Mode</td>
<td>1010</td>
<td>1</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Flash Strength</td>
<td>1011</td>
<td>0/10</td>
<td>signed rational</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Macro</td>
<td>1020</td>
<td>0</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Focus mode</td>
<td>1021</td>
<td>0</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Slow Sync</td>
<td>1030</td>
<td>0</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Picture Mode</td>
<td>1031</td>
<td>1</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Cont. Take Bracket</td>
<td>1100</td>
<td>0</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>unknown</td>
<td>1200</td>
<td>0</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Blur warning</td>
<td>1300</td>
<td>0</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Focus warning</td>
<td>1301</td>
<td>0</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AE warning</td>
<td>1302</td>
<td>0</td>
<td>unsigned short</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

### EXIF Interoperability IFD

<table>
<thead>
<tr>
<th>Entry</th>
<th>Tag</th>
<th>Value</th>
<th>Meaning</th>
<th>Format</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interoperability Index</td>
<td>0001</td>
<td>R98</td>
<td>ascii string</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Interoperability Version</td>
<td>0002</td>
<td>30 31 30 30</td>
<td>undefined</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

### IFD 1

<table>
<thead>
<tr>
<th>Entry</th>
<th>Tag</th>
<th>Value</th>
<th>Meaning</th>
<th>Format</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compression</td>
<td>0103</td>
<td>6</td>
<td>JPEG</td>
<td>unsigned short</td>
<td>1</td>
</tr>
<tr>
<td>Orientation</td>
<td>0112</td>
<td>1</td>
<td>0 degrees</td>
<td>unsigned short</td>
<td>1</td>
</tr>
<tr>
<td>X Resolution</td>
<td>011a</td>
<td>72/1</td>
<td>72.00</td>
<td>unsigned rational</td>
<td>1</td>
</tr>
<tr>
<td>Y Resolution</td>
<td>011b</td>
<td>72/1</td>
<td>72.00</td>
<td>unsigned rational</td>
<td>1</td>
</tr>
<tr>
<td>Resolution Unit</td>
<td>0128</td>
<td>2</td>
<td>inch</td>
<td>unsigned short</td>
<td>1</td>
</tr>
<tr>
<td>JPEG IF Offset</td>
<td>0201</td>
<td>1074</td>
<td></td>
<td>unsigned long</td>
<td>1</td>
</tr>
<tr>
<td>JPEG IF Byte Count</td>
<td>0202</td>
<td>8691</td>
<td></td>
<td>unsigned long</td>
<td>1</td>
</tr>
<tr>
<td>YCbCr Positioning</td>
<td>0213</td>
<td>2</td>
<td>datum point</td>
<td>unsigned short</td>
<td>1</td>
</tr>
</tbody>
</table>
The Dublin Core Metadata Initiative (DCMI) is an organization dedicated to promoting the widespread adoption of interoperable metadata standards. Developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems.
Who can benefit from using Dublin Core metadata?

- Anyone can use Dublin core metadata to describe the resources of an information system
- Web pages are one of the most common types of resources to utilize the Dublin Core's descriptions, usually within HTML’s meta tags; however, increasingly there are many digital archives of physical objects that are starting to make use of the Dublin Core

WHY DUBLIN CORE?

- Provides a great deal of flexibility
- Easy to learn
- Ensures interoperability with other schemes
FIFTEEN OPTIONAL, REPEATABLE ELEMENTS

<table>
<thead>
<tr>
<th>Content</th>
<th>Intellectual Property</th>
<th>Instantiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Creator</td>
<td>Date</td>
</tr>
<tr>
<td>Subject</td>
<td>Publisher</td>
<td>Type</td>
</tr>
<tr>
<td>Description</td>
<td>Contributor</td>
<td>Format</td>
</tr>
<tr>
<td>Source</td>
<td>Rights</td>
<td>Identifier</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SAMPLE DUBLIN CORE RECORD

**SAMPLE RECORD DISPLAY**

<table>
<thead>
<tr>
<th>Identifier</th>
<th>CDS5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Nelle Lutcher</td>
<td></td>
</tr>
<tr>
<td>Interviewer</td>
<td>Willard, Patrica</td>
<td></td>
</tr>
<tr>
<td>Interviewee</td>
<td>Lutcher, Nelle</td>
<td></td>
</tr>
<tr>
<td>Publisher</td>
<td>Rutgers University Institute of Jazz Studies</td>
<td></td>
</tr>
<tr>
<td>Date Recorded</td>
<td>1979-01-10</td>
<td></td>
</tr>
<tr>
<td>Object Type</td>
<td>Sound</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
<td></td>
</tr>
<tr>
<td>Object Format</td>
<td>sound/cd-audio</td>
<td></td>
</tr>
<tr>
<td>Physical Extent</td>
<td>12:17:20:32 KB; 44.2 kHz; 64 Kbps; stereo; 16-bit.</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Lutcher, Nelle—Interviews</td>
<td></td>
</tr>
<tr>
<td>Access Rights</td>
<td>Open Access</td>
<td></td>
</tr>
<tr>
<td>Reproduction Rights</td>
<td>This object may be copyright-protected. Permission to reuse, publish or reproduce this object must be obtained from the publisher or the copyright holder.</td>
<td></td>
</tr>
<tr>
<td>Related Objects</td>
<td>is Format of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.MP3 05 which has the type</td>
<td></td>
</tr>
<tr>
<td></td>
<td>audiorecording consisting of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 audiotapes (reel-to-reel); and has the format</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of 5 audiotapes (reel-to-reel)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>has Format .LT3_02:31:06:47 which has the type</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cd-audio consisting of .LT (digital linear tape)</td>
<td></td>
</tr>
</tbody>
</table>
IDENTIFIER

- Unique identifier assigned to an item; may be an LC class number, an accession number, alphanumeric identifier, etc.
- An alphanumeric accession number will be used for the project.
- example- DAT199 = Digital Audio Tape 199.

CREATOR

- Person or body primarily responsible for content of the resource.
- Qualified by “Role” to further define Creator.
- “Interviewer” is used in the qualifier Role for the project.
TITLE

- Name given to a resource; typically a name by which a resource is formally known.
- Qualifier of “Alternative” is used to provide title variations.
- example- Title: Charles Mingus (Nat Hentoff)
  Alternative: Nat Hentoff interviews Charles Mingus

How is Dublin Core metadata stored?

- Dublin Core metadata is often stored as name-value pairs within META tags, which are placed within the HEAD elements of an HTML document
  
  - However, it can also be located in an external document or loaded into a database enabling it to be indexed and manipulated from within a propriety application
    - Elements stored in a relational database, using SQL
Use of Dublin Core

- educational organizations
- libraries
- government institutions
- scientific research sector
- Web page authors
- businesses requiring more searchable sites
- corporations with vast knowledge management systems

Resource Description Framework (RDF)

- World-Wide Web Consortium is developing a standard for metadata
- RDF foundation to provide processing metadata in web documents
- Preferred language for writing RDF schemas is XML
XML as a Common Syntax

- XML (and SGML) provide a way of expressing the structure of documents that can be verified and validated by document processing systems.
- XML thus provides a way of representing metadata descriptions as well as the content that they describe.

What are XML Schemas?

- An XML vocabulary for expressing your data’s structure AND content types, and even the business rules involved in processing the data.
- Written in XML themselves.
- Support namespaces for combining multiple schemas in the same documents.
Are all our problems solved?

- "...the formal release of MPEG-7, MPEG-21, 2001, 2002, 2003 and 2004 respectively brought an end to the 1990s struggles with indexing content...."
- "Multimedia indexing approaches and subsequent instalments are an important historical milestone but they represent the previous generation of indexing techniques"
- "Since September 2001, the multimedia community has had to re-engineer their schemes, languages, navigation tools to work with at least one of the open MPEG standards"

MPEG Standards

- MPEG-7
  - Describing the multimedia content data that supports some degree of interpretation of the information’s meaning, which can be passed onto, or accessed by, a device or a computer code
- MPEG-21
  - A normative open framework for multimedia delivery and consumption for use by all the players in the delivery and consumption chain
MPEG-7 Motivation

- Create standardized multimedia description framework
- Enable content-based access to and processing of multimedia information on the basis of descriptions of multimedia content and structure (metadata)
- Support range of abstraction levels for metadata from low-level signal characteristics to high-level semantic information

MPEG-7 Query Examples

- Play a few notes on a keyboard and retrieve a list of musical pieces similar to the required tune, or images matching the notes in a certain way, e.g., in terms of emotions
- Draw a few lines on a screen and find a set of images containing similar graphics, logos, ideograms, ...
- Define objects, including color patches or textures and retrieve examples among which you select the interesting objects to compose your design

- So all the problems are solved?
http://www.chiariglione.org/mpeg/
MPEG-7 Metadata Framework

- **Data**
  - “multimedia information that will be described using MPEG-7, regardless of storage, coding, display, transmission, medium, or technology.”

- **Feature**
  - “a distinctive characteristic of the data [that] signifies something to somebody.”

MPEG-7 Description Tools
MPEG-7 Top Level Hierarchy

MPEG-7 Still Image Description
MPEG-7 Video Segments Example

MPEG-7 Segment Relationship Graph
MPEG-7 Conceptual Description

MPEG-7 Summaries
MPEG-7 Application Framework

- MPEG-7 based search engines will simply need to process values of defined standard parameters
- The do not need to compute any search features
Many years ago there lived an emperor who was quite an average fairy tale ruler, with one exception: he cared much about his clothes.

One day he heard from two swindlers named Guido and Luigi Farabutto that they could make the finest suit of clothes from the most beautiful cloth.

This cloth, they said, also had the special capability that it was invisible to anyone who was either stupid or not fit for his position.
Being a bit nervous about whether he himself would be able to see the cloth, the emperor first sent two of his trusted men to see it.

Of course, neither would admit that they could not see the cloth and so praised it.

All the townspeople had also heard of the cloth and were interested to learn how stupid their neighbors were.

The emperor then allowed himself to be dressed in the clothes for a procession through town, never admitting that he was too unfit and stupid to see what he was wearing. For he was afraid that the other people would think that he was stupid.

Of course, all the townspeople wildly praised the magnificent clothes of the emperor, afraid to admit that they could not see them, until a small child said:

"But he has nothing on!"
It is important to remember that MPEG-7 does not offer any solution for the problems of feature extraction and matching!!!

These issues will remain within the research domain in the future

The semantic gap will remain as wide as ever!!!

MPEG-7 Applications Today...

- IBM MPEG-7 Annotation Tool
  - Assists in annotating video sequences with MPEG-7 metadata
- Ricoh MPEG-7 MovieTool
  - A tool for creating video content descriptions conforming to MPEG-7 syntax interactively
- Canon MPEG-7 Speech Recognition engine
  - Web site allows you to create an MPEG-7 Audio “SpokenContent” description file from an audio file in “wav” format
IBM MPEG-7 Annotation Tool

- The IBM MPEG-7 Annotation Tool assists in annotating video sequences with MPEG-7 metadata
  - Each shot in the video sequence can be annotated with static scene descriptions, key object descriptions, event descriptions, and other lexicon sets
  - The annotated descriptions are associated with each video shot and are stored as MPEG-7 descriptions in an XML file
  - Can also open MPEG-7 files in order to display the annotations for the corresponding video sequence
  - Customized lexicons can be created, saved, downloaded, and updated
Ricoh MovieTool

- Creates an MPEG-7 description by loading video data
- Provides visual clues to aid the user in creating the structure of the video
- Automatically reflects the structure in the MPEG-7 descriptions
- Visually shows the relationship between the structure and MPEG-7 descriptions
- Presents candidate tags to help choose appropriate MPEG-7 tags
- Checks the validation of the MPEG-7 descriptions in accordance with MPEG-7 schema
- Can describe all metadata defined in MPEG-7
- Is able to reflect any future changes and extensions made to MPEG-7 schema

Canon MPEG-7 ASR Tool

Welcome to the MPEG-7 Spoken Content Transcription service

This web site allows you to create an MPEG-7 SpokenContent description file from an audio file in .wav format. The software is based on speech recognition technology being developed by Canon.

MPEG-7 is a new standard designed to facilitate the archiving, searching and browsing of multimedia content in the web 2.0 era. The audio part of the standard, known as ASR (Automatic Speech Recognition), is still under development. This early prototype of the software allows even if word recognition rates are relatively low due to small vocabulary and high background noise levels.

© Copyright 2003 Canon Research Centre Europe
MPEG-7 Future

- New application specific profiles
- Integration into media production and reuse cycle
  - Automated metadata creation in devices
    - Automatic annotation (example: digital cameras - EXIF!)
    - Example, time, color histogram
  - Use of MPEG-7 metadata in multimedia applications
- MPEG-21

means of sharing digital rights/permissions/restrictions for digital content from content creator to content consumer

![Diagram of MPEG-7 Future Structure]

- Resources
  - MPEG-1
  - MPEG-2
  - MPEG-4
- Meta-Data
  - MPEG-7
- New Meta-Data and Resource Standards

MPEG-21
Second problem....

- MPEG-7 standard does not define how the searching or indexing has to be made
  - Indexing will remain within the research domain in the future

- It also does not make any assumptions about the internal storage format in a database

- Semantic Gap
- Manual textual annotation
  - Subjective process
- Meta data generation
- MP3
- EXIF
- Dublin Core
- MPEG-7
  - MPEG-7 and the semantic gap
  - MPEG-7 Annotation
  - MPEG future, MPEG-21
  - MPEG-7 and indexing