5 Multimedia Content Description

Literature:
Rosenblatt/Trippe/Mooney, Digital Rights Management, Chapter 6

Unlabelled Video Tapes & The Internet

- The Unlabelled Video Tape Problem
  - Even worse with digital media: Various formats, variants
- Digital media production:
  - Labelling of parts to be composed
    - » Date, time, format, ...
  - Representing the composition
- Digital media on the Internet
  - Identifying digital media
    - » Title, author, genre, ...
  - Searching for specific media, e.g. audio, video content
  - Fine-grained search within media
    - » e.g. person search within video content
  - Bringing together related media (e.g. text news and photos)
    - » (Automated) syndication

Content, Essence, Metadata

- Content
  - consists of *essence* data and *metadata*
- Essence
  - parts of content that directly represent program material such as audio, video, graphic, still-image, text, or sensor-data
- Metadata
  - parts of content that contain data used
    - » to *describe* essence or
    - » to provide information on its *use*
  - metadata objects sometimes called “mobs”
- Metadata may be
  - Stored separately from the essence data
  - Combined with the essence data (“embedded metadata”)

Source: AAF Developer Overview

Metadata Problems

- Creation metadata
  - During the creation of media essence, metadata is created but often ignored
  - Example: EXIF data in JPEG
- Manually added metadata
  - Users notoriously ignore the administration of metadata
- Metadata incompatibility
  - Metadata exists in various formats specific for media types, applications, product vendors, ...
  - Exchange of metadata is difficult
- Broad range of metadata
  - Metadata exists on various levels, covering all is expensive
- Metadata economy
  - How much of the metadata will be used?
  - When to create metadata?

Selected Metadata Standards

- Dublin Core Metadata Initiative (DCMI) &
  PRISM (Publishing Requirements for Industry Standard Metadata)
  - Oriented towards books, magazines, journals etc.
  - Uses XML, RDF, Dublin Core
  - www.prismstandard.org
- ONIX (Online Information Exchange)
  - For books: http://www.editeur.org/onix.html
- MUZE (www.muze.com)
  - De-facto industry standard
  - Company collecting large database of recorded music
- TV Anytime (www.tv-anytime.org)
  - Devoted to audio-visual services making use of local mass-storage
  - Focus on: Electronic Program Guide and user profiles
- EBU P/Meta
  - Devoted to material exchange between broadcasting stations

Dublin Core Metadata Initiative DCMI

- See http://dublincore.org
- DC Simple level:
  - Title, Subject, Description, Type, Source, Relation, Coverage, Creator, Publisher, Contributor, Rights, Date, Format, Identifier, Language
- DC Qualified level:
  - Audience, Provenance, RightsHolder, semantic refinements
- Principles:
  - One-to-one: Decribes only one manifestation of a resource
  - Dumb-down: Element+value shall be interpretable even when ignoring existing qualifiers

5 Multimedia Content Description

5.1 Metadata: Concepts and Overview

5.2 RDF: XML Metadata

5.3 Metadata for Authoring: AAF & SMPTE Standards

5.4 Generic Metadata Framework: MPEG-7

(continued)

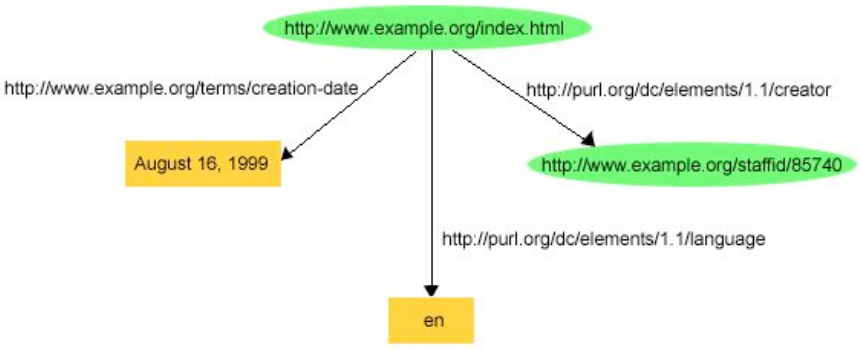
Literature:
www.w3c.org/RDF

Resource Description Framework RDF

- Language for representing information about resources in the WWW
  - W3C's Semantic Web activity
- *Resource:* Anything that can be identified by a URI (e.g. all Web pages)
- *Property:* An attribute of a described thing which can take on specific values
- *Statement:* A triple consisting of
  - *Subject:* Some resource to be described
  - *Predicate:* A property of the subject
  - *Object:* A specified value
- Properties, values and statements are resources themselves,
  - i.e. can be identified by a URI
  - i.e. can be subject to further description

RDF Example

- **http://www.example.org/index.html** has a **creator** whose value is **John Smith**
- **http://www.example.org/index.html** has a **creation-date** whose value is **August 16, 1999**
- **http://www.example.org/index.html** has a **language** whose value is **English**



RDF/XML Example

- RDF/XML is an XML language for representing descriptions

```
<?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/1.1/"
    xmlns:exterms=
      "http://www.example.org/terms/">
  <rdf:Description
      rdf:about="http://www.example.org/index.html">
    <exterms:creation-date>August 16,1999
      </exterms:creation-date>
    <dc:language>en</dc:language>
    <dc:creator
      rdf:resource="http://www.example.org/staffid/85740"/>
  </rdf:Description>
</rdf:RDF>
```

Example: Audio Metadata in DC-based RDF/XML

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description
      rdf:about="http:www.medien.ifi.lmu.de/team/
      heinrich.hussmann/files/mmn5a.m4b>
    <dc:creator>Heinrich Hussmann</dc:creator>
    <dc:title>Multimedia Content Description I</dc:title>
    <dc:description>Discusses multimedia metadata
      standards.</dc:description>
    <dc:date>2006-11-24</dc:date>
    <dc:format>audio/mp4</dc:format>
  </rdf:Description>
</rdf:RDF>
```


Example RDF Application: Music Metadata

- **NO official standard!**
- MusicBrainz Metadata Initiative (http://musicbrainz.org/MM/)
- RDF/XML-based standard for audio/video related metadata
  - To be combined with Dublin Core metadata recommendation
- Application: Queries and results for music metadata
  - Similar to CDDB but much more flexible
  - Database contents created by community
- Application: Automatic tagging of existing music files with metadata
  - Based on acoustic fingerprints
    - » TRM (TRM Recognizes Music)
    - » PUID (Portable Unique Identifier) from MusicDNS fingerprinting
  - “Tagger” software writes MP3 ID3 tags

MusicBrainz RDF Example (1)

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc  = "http://purl.org/dc/elements/1.1/"
  xmlns:mm  = "http://musicbrainz.org/mm/mm-2.1#">
<mm:Album
      rdf:about="http://musicbrainz.org/album/1d16698f-e1fc-
      48b3-ab4a-941c2ecec5bb">
  <dc:title>Rubycon</dc:title>
  <dc:creator rdf:resource="http://musicbrainz.org/artist/
      23d8426c-18c7-46e6-a51d-7395bd43c641"/>
  <mm:cdindexidList>
      <rdf:Bag>
           <rdf:li rdf:resource="http://musicbrainz.org/
                cdindex/borOdvYNUkc2SF8GrzPepad0H3M-"/>
       </rdf:Bag>
  </mm:cdindexidList>
  <mm:releaseType rdf:resource="http://musicbrainz.org/mm
           /mm-2.1#TypeAlbum"/>
  <mm:releaseStatus rdf:resource="http://musicbrainz.org/mm/
           mm-2.1#StatusOfficial"/>
...
```


MusicBrainz RDF Example (2)

```
...
  <mm:trackList>
    <rdf:Seq>
      <rdf:li rdf:resource="http://musicbrainz.org/track/
           82e7dbbd-814a-4c46-bd0c-d7d65af67a3f"/>
      <rdf:li rdf:resource="http://musicbrainz.org/track/
           d94fc569-96b3-4c75-abd5-d0d381d7cc58"/>
    </rdf:Seq>
  </mm:trackList>
</mm:Album>

<mm:Artist rdf:about="http://musicbrainz.org/artist/
           23d8426c-18c7-46e6-a51d-7395bd43c641">
  <dc:title>Tangerine Dream</dc:title>
  <mm:sortName>Tangerine Dream</mm:sortName>
</mm:Artist>

<mm:Track ... Mm:Duration ...
</mm:Track>
</rdf:RDF>
```

Example: MusicVocabulary

- Unofficial proposal for a metadata schema describing classical music and performances
  - http://www.kanzaki.com/ns/music
- Defined in OWL-DL (extension of RDF)
- Example:

```
<m:Concert>
  <dc:title>Tokyo Green Symphony Orchestra 12th
    Concert</dc:title>
  <dc:date>2003-11-02T14:00+09:00</dc:date>
  <m:conductorName>Yuri Nitta</m:conductorName>
  <m:performerName>Violin: Tomoko Kawada</m:performerName>
  <m:performerName>Orchestra: Tokyo Green Symphony
    Orchestra</m:performerName>
  <m:programLine>
    Bruch: Violin Concerto No.1 g minor</m:programLine>
  <m:programLine>Brahms: Symphony No.2 D major
  </m:programLine>
  <cal:location>Sumida Triphony</cal:location>
</m:Concert>
```


5 Multimedia Content Description

Literature:

AAF Developer Overview, available at www.aafassociation.org

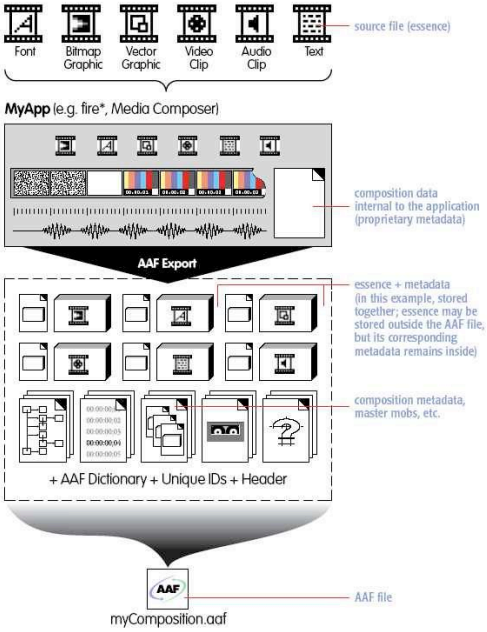
Integration of Digital Media in Video Production

- Example: Putting together all audio elements for a film soundtrack
  - music tracks, ambient sound tracks, performer's synchronized sound, ...
- Multimedia authoring applications
  - Usually use proprietary data formats
  - Important metadata related to creation process (e.g. compositional metadata) kept only in proprietary formats
- Standards in the broadcasting industry
  - SMPTE (Society of Motion Picture and Television Engineers)
  - EBU (European Broadcasting Union)
  - Working on hardware-based standards for a long time
- EBU/SMPTE Task Force for Harmonized Standards for the Exchange of Program Material as Bit Streams (1996-1999)
  - Results further developed into Advanced Authoring Format (AAF)
  - AAF: Industry-driven, cross-platform, multimedia file format

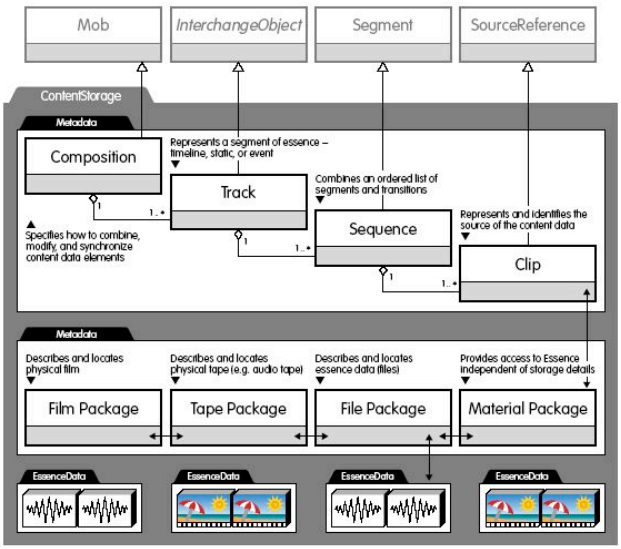
Types of Metadata Covered in AAF

- **Identification and Location Metadata**
  - comprises all forms of metadata that can be used to uniquely identify an item
- **Administration Metadata**
  - definitions of rights, user access, security classifications, encryption, audience listings and other business information.
- **Interpretive Metadata**
  - partly for human-orientated metadata types such as names, artists, organisations and classification.
- **Parametric Metadata**
  - signal coding parameters, device characteristics, sensor parameters (e.g.focal length) plus device storage and streaming parameters
- **Process Metadata**
  - includes all items that describe how essence is assembled, such as editing and compositional metadata
- **Relational Metadata**
  - describes how information is related
- **Spatio-Temporal Metadata**
  - describes places and time including angles, geo-spatial coordinates, dates, creation times, event times, delays and durations

Interchanging Compositions with AAF



AAF Object-Oriented Software Architecture



5 Multimedia Content Description

5.1 Metadata: Concepts and Overview

5.2 RDF: XML Metadata

5.3 Metadata for Authoring: AAF & SMPTE Standards

5.4 Generic Metadata Framework: MPEG-7

(continued)

Literature:

www.chiariglione.org

B. S. Manjunath, Philippe Salembier, Thomas Sikora:
Introduction to MPEG-7, Wiley 2002

MPEG-7

- Moving Picture Experts Group (MPEG)
  = ISO/IEC JTC1/SC29/WG11 “Moving Pictures and Audio”
  Main Web presence now: www.chiariglione.org
- MPEG-7 “Multimedia Content Description Interface”
  - “ ... a standard for 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-7 is not aimed at any one application in particular; rather, the elements that MPEG-7 standardizes support as broad a range of applications as possible.”
- Version 1 developed in 1996 – 2001
- Version 2 said to be under development
- Industrial uptake rather slow
  - Very ambitious standard
- Some research and open source prototypes available
  - See the Wikipedia links for MPEG-7

Parts of the MPEG-7 Standard

- MPEG-7 Systems
- MPEG-7 Description Definition Language (DDL)
  - Descriptors (D) define the syntax and semantics of each *feature* (metadata element)
  - Description schemes (DS) specify syntax and semantics of the relationships between their components, which may be Descriptors or Description Schemes
  - DDL allows the creation of Ds and DSs
    - » XML-based language with some small extensions to XML Schema
- MPEG-7 Visual
- MPEG-7 Audio
- MPEG-7 Multimedia Description Schemes
- MPEG-7 Reference Software
  - eXperimentation Model XM

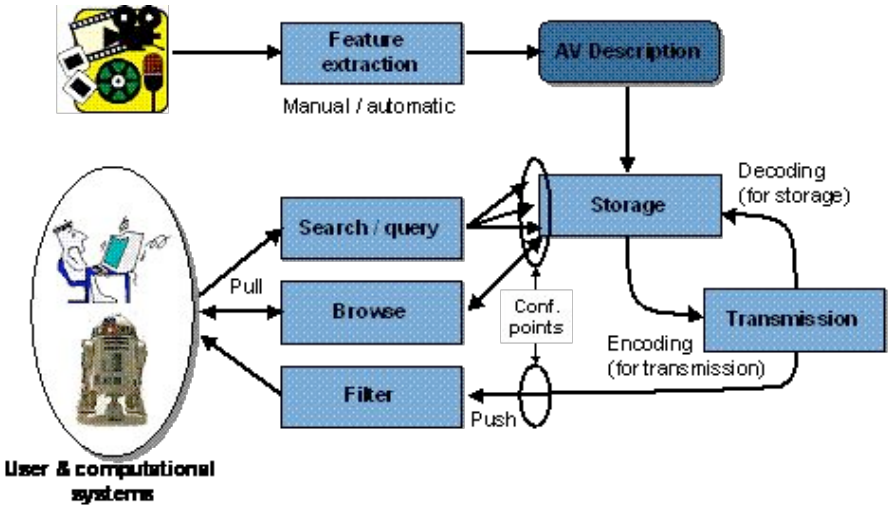
Types of Metadata Covered in MPEG-7

- Technical Metadata:
  - Form (data format, representation parameters like resolution, colour depth...)
  - For live captured material: Time and date of original occurrence
  - Technical parameters of capture (e.g. aperture, exposure etc. for images)
- Content Description Metadata (main focus of MPEG-7):
  - Low level:
    - » Video: Shapes, positions, trajectories etc. of objects ("an object with mainly yellow colour fitting into a box moving from left to right")
    - » Audio: Key, mood, tempo, tempo changes, position in sound space, ...
  - High level:
    - » Video: "A post car arrives, entering the scene from the left side."
    - » Audio: Title, composer, structure etc. or, e.g.: "barking dog"
- Additional information:
  - Digital rights, classification, context, further links, ...

Application Areas of MPEG-7

- Architecture, real estate, and interior design (e.g., searching for ideas).
- Broadcast media selection (e.g., radio channel, TV channel).
- Cultural services (e.g., virtual museums).
- Digital libraries (e.g., image catalogue, musical dictionary).
- Education (e.g., repositories of multimedia courses).
- Home Entertainment (e.g., home video management).
- Investigation services (e.g., human characteristics recognition, forensics).
- Journalism (e.g. searching for video footage of political event).
- Multimedia directory services (e.g. yellow pages, Tourist information).
- Multimedia editing (e.g., personalized electronic news service, media authoring).
- Remote sensing (e.g., cartography, ecology, natural resources management).
- Shopping (e.g., searching for clothes that you like).
- Surveillance (e.g., traffic control, surface transportation).
- ...

A Hypothetical MPEG-7 Chain

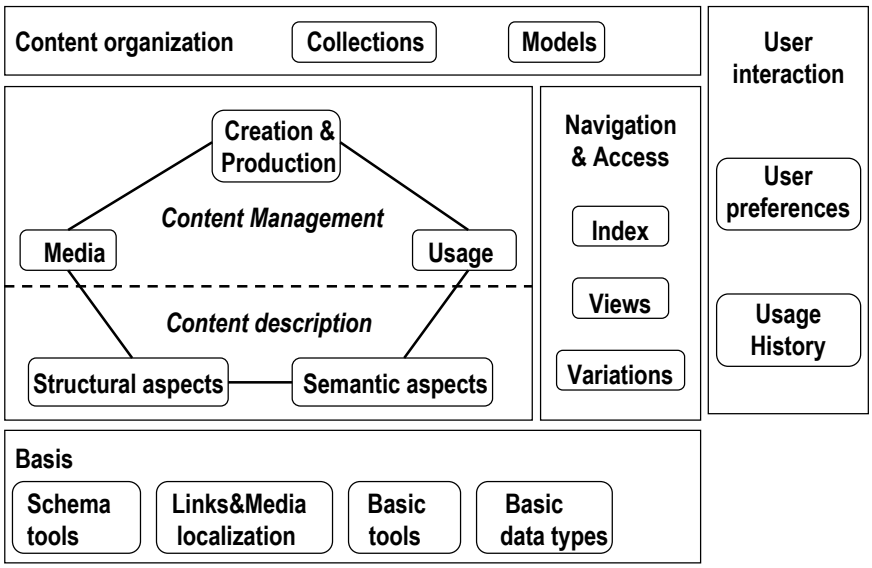


Examples of Advanced Queries

- 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 colour patches or textures and retrieve examples among which you select the interesting objects to compose your design.
- On a given set of multimedia objects, describe movements and relations between objects and so search for animations fulfilling the described temporal and spatial relations.
- Describe actions and get a list of scenarios containing such actions.
- Using an excerpt of Pavarotti's voice, obtaining a list of Pavarotti's records, video clips where Pavarotti is singing and photographic material portraying Pavarotti.

From: MPEG-7 Overview

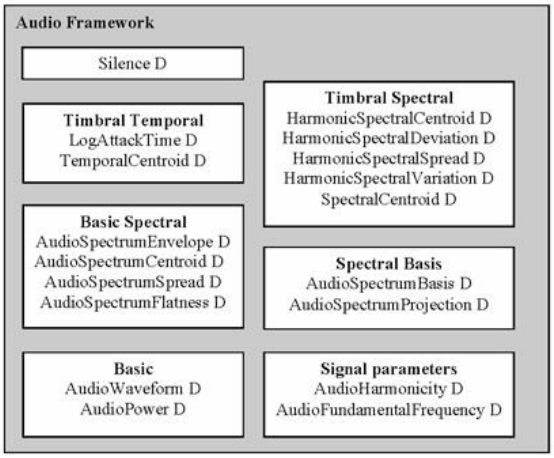
Organization of Multimedia Description Tools



MPEG-7 Audio Low-Level Descriptors

- Structures:
  - Single scalar value
  - Series of sampled values
- Features:
  - See figure

(More details see next lecture)



MPEG-7 Audio High-Level Descriptors

- Audio signature
  - Statistical summary of spectral flatness descriptor
  - Fingerprinting, identification of audio content
- Musical instrument timbre
- Melody description
  - MelodyContour (terse, efficient)
  - MelodySequence
    - » Query by Humming
    - » Example: http://www.musicline.de/ --> Melodiesuche
- General sound recognition and indexing
  - Probabilistic classifiers for sound classes
- Spoken content
  - Output and intermediate results of Automatic Speech Recognition (ASR)

Example: Low-Level Audio Feature Extraction

- TU Berlin MPEG-7 Audio Analyzer
  - See http://mpeg7lld.nue.tu-berlin.de/
- Demo:
  - Uploaded file with cello sample
  - Request feature extraction
  - Download XML file with feature description
- Observation:
  - Extremely low level: Just numbers from spectral analysis
  - Requires sophisticated post-processing

Demo MPEG-7 Feature Description (1)

```
<?xml version="1.0" encoding="iso-8859-1"?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
  xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-
  2001.xsd">
<Description xsi:type="ContentEntityType">
<MultimediaContent xsi:type="AudioType">
<Audio xsi:type="AudioSegmentType">

<AudioDescriptor xsi:type="AudioWaveformType"  minRange="-
  0.773438" maxRange="0.843750">
  <SeriesOfScalar hopSize="PT10N1000F"
  totalNumOfSamples="170" >
      <Min> -0.312500 -0.304688 -0.554688 -0.343750 -0.500000
  -0.468750 -0.468750 -0.562500 -0.539063 ... 0.000000</Min>
      <Max> 0.257813 0.312500 ... 0.000000</Max>
  </SeriesOfScalar>
</AudioDescriptor>

...
```

Demo MPEG-7 Feature Description (2)

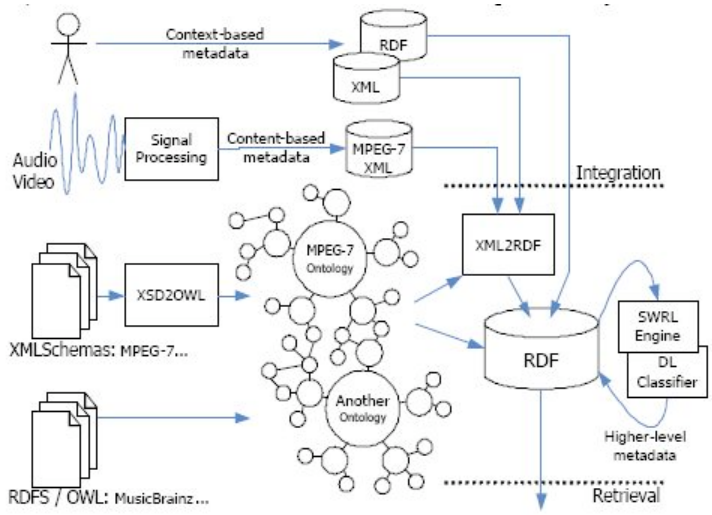
```
<AudioDescriptor xsi:type="AudioPowerType">
  <SeriesOfScalar hopSize="PT10N1000F"
  totalNumOfSamples="170" >
      <Mean> 0.0159396 0.0223927 0.0454712 0.0329396 0.0512068
  0.0399808 ... 1.72008e-005 0</Mean>
  </SeriesOfScalar>
</AudioDescriptor>
...
<AudioDescriptor xsi:type="LogAttackTimeType">
      <Scalar>  -0.130768</Scalar>
</AudioDescriptor>

<AudioDescriptor xsi:type="TemporalCentroidType">
      <Scalar>  0.792369</Scalar>
</AudioDescriptor>
...
</Audio>
</MultimediaContent>
</Description>
</Mpeg7>
```


Embedding MPEG-7 into the Semantic Web

- Semantic Web
  - Making the Web better searchable by including semantic descriptions
  - Basis for information storage: RDF
  - Networks of semantic definitions *(ontologies)* described in OWL (Web Ontology Language)
- Recent research on mapping MPEG-7 metadata terminology onto ontologies:
  - Jane Hunter, Adding Multimedia to the Semantic Web – Building an MPEG-7 Ontology, Proceedings of the First Semantic Web Working Symposium (SWWS), Stanford, USA (2001) 261-281
  - Roberto Garcia, Oscar Celma, Semantic Integration and Retrieval of Multimedia Metadata, Proceedings of 4th International Semantic Web Conference, Galway, Ireland (2005)
    - » http://www.iua.upf.edu/mtg/publications/d450b9-ISWC2005-GarciaCelma-SemAnnot2005.pdf

MPEG-7 Metadata Integration Architecture



Metadata integration and retrieval architecture

Garcia/Selma

Semantic Integration of Different Terminologies

MusicBrainz to MPEG-7 OWL ontology mappings

```
musicbrainz:Artist ⊆ mpeg7:CreatorType
musicbrainz:Album ⊆ mpeg7:CollectionType
musicbrainz:Track ⊆ mpeg7:AudioSegmentType
dc:author ⊆ mpeg7:Creator
dc:title ⊆ mpeg7:Title
musicbrainz:sortName ⊆ mpeg7:Name
musicbrainz:duration = mpeg7:MediaDuration
```

Music Vocabulary ontology to MPEG-7 OWL ontology mappings

```
music:Music_Unit ⊆ mpeg7:AudioSegmentType
music:sections = mpeg7:AudioSegment
music:Artist ⊆ mpeg7:CreatorType
```

Garcia/Selma