

Introduction to MPEG-7 and its Applications



Artur Lugmayr

lartur@acm.org

www.lugy.org

'Artturi'

Slides can be found on:

www.digitalbroadcastitem.tv

As well as a few figures are excerpted from:

A. Lugmayr, S. Niiranen, and S. Kalli, "Digital Interactive TV and Metadata", Springer-Verlag 2004, NY

Table of Content

- A bit about myself...
- Problem definition
- Demonstrations
- Metadata
- MPEG-7 General
- MPEG-7 Specific
- Case studies

Artur Lugmayr



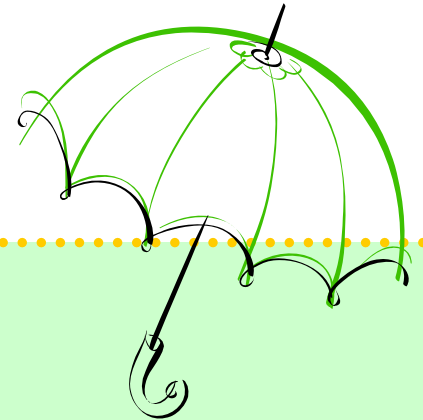
- World's first book "Digital interactive TV and metadata – Future broadcast multimedia", Springer-NY, 2004, www.digitalbroadcastitem.tv
- **Current Position**
 - Team leader: DigitalTV/NAMU research group at the Digital Media Institute/TUT, Tampere, Finland
- **Interests**
 - Future oriented immersive entertainment experiences
 - Ambient *smart* media
 - New concepts in multimedia
- **Professional**
 - National/international projects
 - Engel Virtual Machine (EVM), omVR, futureTV, future interaction TV, Umedia (EU, NoE), etc.
- **Activities**
 - ISO/IEC MPEG standardization
 - Previously chairman of MPEG-21 in broadcasting
- **Inventions**
 - Digital Broadcast Item Model (DBIM)
- **Scientific**
 - **Book + book chapter**
 - **>15 scientific publications**
 - **4 MSc. thesis supervision**
- **Why am I here???**

What IS the problem?

- **'Myriads' of home-devices**
 - 96% possess a TV, 98.8% a radio, 72% a VCR, 63% a wireless phone, 54.2% a wired phone, 49% a PC, 40% Internet access in the EU...
- **Many distribution channels + protocols**
 - Terrestrial, cable, satellite, wireless, INTERNET (!), IP, TCP/IP, SOAP, ...
- **Interactive consumers**
 - According to studies 44% of the consumers in the US actively contribute to online content via posted photographs, accessing files, written material, websites, etc.
 - 24/7 interactivity, content synchronized interactive services, inner-content alternation, content related services
 - 60% of users used the Internet for buying, information services, job-related, fun, education, communication and for searching answers to questions
- **Digital switchover in broadcasting**
 - The digital switchover from analogue to digital TV is the next huge development in broadcasting after the introduction of colour TV in 1960 and *should* be realized until 2006
 - 59% of turnovers in the audio-visual market in the year 2000 were attributed to TV broadcasting in the EU

There are already, or will be...

millions of content provider
millions of broadcasters
millions of service provider
millions of aggregators
millions of users
millions of different formats
millions of devices

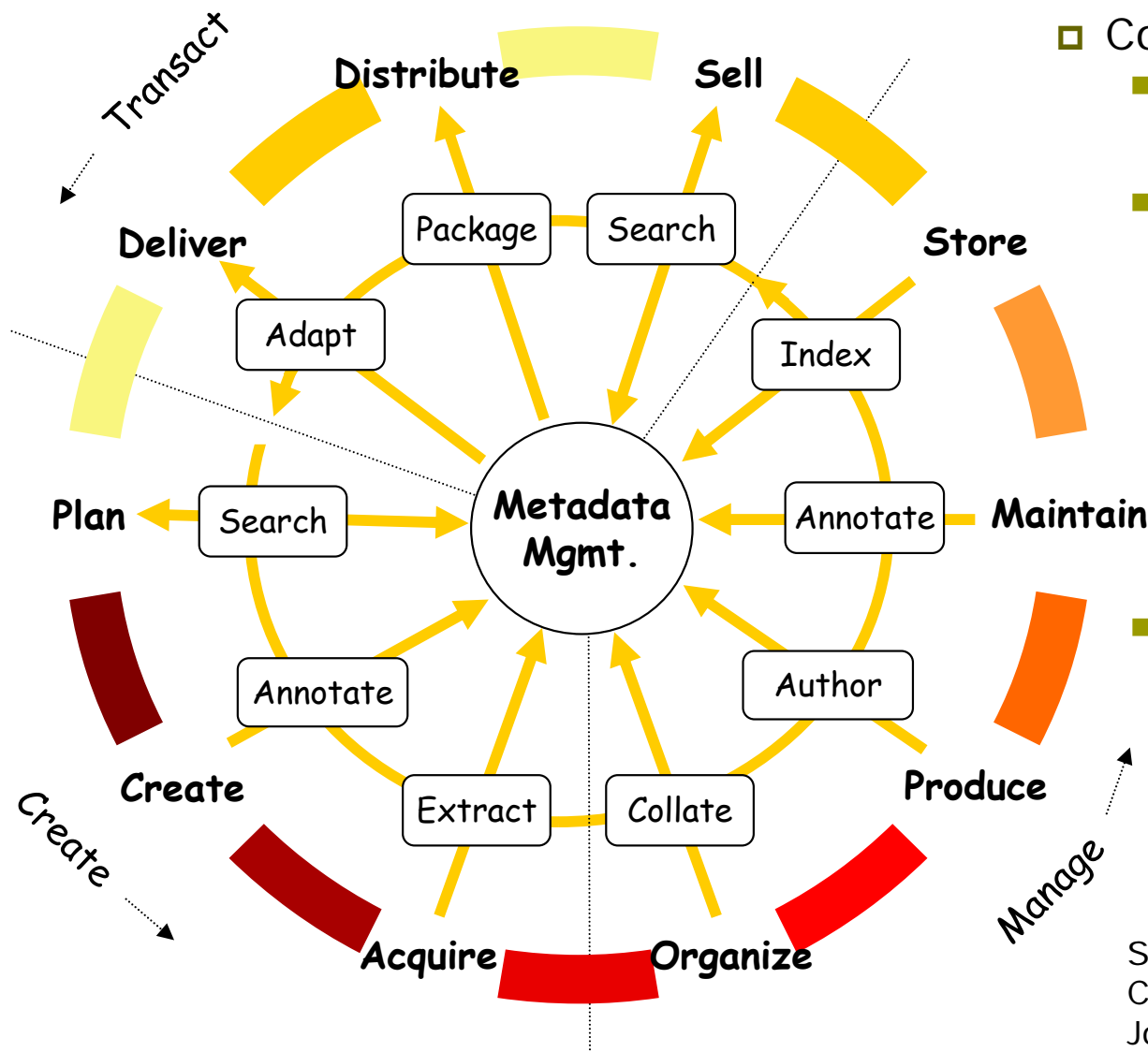


CONCLUSION

There is a need to provide broadcasters with...

- a **value-chain** spanning framework
- for **transparent** and augmented delivery
- and **use** of digital services
- over **any type** of *distribution channel*,
- to any type of *consumer device*,
- **packaging** solution of digital content,
- **interoperability**,
- enabling **interactivity**
- **by minimizing production costs...**

Digital Media Value Chain (MPEG-7 Metadata Management)



Content value chain:

- Life-cycle:
 - Increase content asset value through repurposing
- Metadata layer:
 - First-class role of metadata (smart bits) throughout digital media lifecycle
 - Central role of metadata management (XML schemas, catalog records, ontologies)
 - MPEG-7 for content description
 - MPEG-21 for packaging, rights mgmt, transactions
- Digital media metadata access functions:
 - **Create:** search, annotate, extract
 - **Manage:** index, annotate, collate
 - **Transact/Distribute:** search, package, adapt

Source: MPEG-7 Multimedia Content Description Standard, John R. Smith, IBM, Presentation online, January 8, 2003

Current issues in Multimedia

- Interoperability
- Seamless integration of devices
- Management of 'millions'
- Adaptive & scaleable content
- Need for packaging of content
- Digitalization and automation of the value-chain
- Adding 'interactivity' to broadcast content
- Broadcast content data management
- Framework for packaging and distribution
- Digital rights management (DRM)
- Quality of Service (QoS)
- ...

**Metadata IS the GLUE
to 'manage millions'!!**

What is Metadata???

- "Data about Data"
- Metadata standards are predefined data structures for describing data to support interoperability, and unified services
- **Objective**
 - factual information e.g. programme name
- **Subjective**
 - attributes e.g. variable information (time)
- **XML is the key**
 - XML Schemas, -Instances
- **Example:**
 - Programme Schedule
 - Imaging Standards
 - Date: 3rd April 2002

Ma Ti Ke To Pe La Su
[01](#) [02](#) [03](#) [04](#) [05](#) [06](#) [07](#)

TV2
Keskiviikko 3.4.2002

5.55 – 9.15 **Uutisikkuna**

12.25 **YLE Teeman ohjelmi**
Tiedefoorumi.

13.20 **Ajankohtainen kakko**
Terveystietoja rahal
Englannin malliin - kuk
... vuotta. Hittibi
TV2/Ajankoh



Isku-Volley juhlii lentopallomestaruutta
(Pvm: 01.04.2002, klo: 20.07)

Tamperealainen Isku-Volley on voittanut lentopalloliigan mestaruuden. Isku peittosi Rovaniemellä Perungan Pojat erin 3-0 (25-17,25-23,25-19) ja otti SM-kultaa otteluvoitoin 3-1. Edellisen kerran Isku voitti SM:n keväällä 1988 - tällöin nimellä Rantaperkiön Isku.



tsikoissa
1 ajankohtais

ainen
takana. Länr

Mestareiden liigan puolivälierät käyntiin
(Pvm: 01.04.2002, klo: 20.40)

Jalkapallon Mestarien liigan puolivälierät alkavat tiistaina kahdella huippuottelulla



Where does Metadata come from?

□ Sources

- Metadata Extractors
- by "Hand" – the hard way...
- Legacy Material

□ Standards

- SMPTE
<http://www.smpte.org/>
- TV-Anytime
<http://www.tv-anytime.org/>
- MPEG-7/-21 <http://mpeg.telecomitalia.com/>
- W3C Defined
<http://www.w3c.org/>
 - RDF (Resource Description Framework)
 - XML Definitions
XML Schema, XPath, XLink, SOAP, TVWeb, Semantic Web, CSS, etc.



What is this?

Formal Language Theory

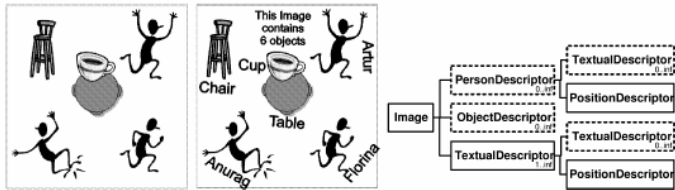
$\Sigma = \{Image, ObjectDescriptor, TextualDescription, PositionDescription, PersonDescription, \#PCDATA\}$
 $V_t = \{\#pcdata\}$
 $V_n = \{Image, ObjectDescriptor, TextualDescriptor, PositionDescriptor, PersonDescriptor\}$
 $S = \{Image\}$
 $R = \{Image \rightarrow PersonDescriptor * ObjectDescriptor * TextualDescriptor +, PersonDescriptor \rightarrow TextualDescriptor * PositionDescriptor * \text{int,y:int} \rightarrow \epsilon, TextualDescriptor \rightarrow \#pcdata\}$

XML Schema

```

<xs:complexType>
  <xs:sequence>
    <xs:element ref="PersonDescriptor" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="ObjectDescriptor" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="TextualDescriptor" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="ID" type="xs:string" use="required"/>
</xs:complexType>

```



(a) Original sample image (Asset Tier). (b) Annotated sample image (Processing Tier). (c) Metadata schema (Description Tier).



(d) Metadata file (Description Tier). (e) Transformed XML file with XSLT as description presentation (Description Tier).



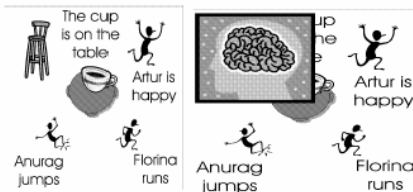
(f) Downloading and linking assets by e.g. Hyperlinks (Linkage Tier).

DTD

```

<!ELEMENT Image (ObjectDescriptor | PersonDescriptor | PositionDescriptor | TextualDescriptor) #IMPLIED
ref (ObjectDescriptor | PersonDescriptor | PositionDescriptor | TextualDescriptor) #IMPLIED
minOccurs CDATA #IMPLIED
maxOccurs CDATA #IMPLIED
type (PositionDescriptionType | xs:string) #IMPLIED
>
<!ELEMENT xs:schema (xs:element | xs:complexType)+>

```



(g) Adding meaning to the asset (Semantic Tier). (h) Conclusions, and implications (Knowledge Tier).

Figure 3.3: This figure shows the process for a metadata system during its life

XML File

```

<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet type="text/xsl" href="\Examples\SimpleImageXSLT.xslt"?>
<Image xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="\Examples\SimpleImageSchema.xsd" ID="10">
  <PersonDescriptor>
    <TextualDescriptor>Anurag</TextualDescriptor>
    <PositionDescriptor x="25" y="25"/>
  </PersonDescriptor>
  <PersonDescriptor>
    <TextualDescriptor>Artur</TextualDescriptor>
    <PositionDescriptor x="75" y="75"/>
  </PersonDescriptor>
  <PersonDescriptor>
    <TextualDescriptor>Florina</TextualDescriptor>
    <PositionDescriptor x="75" y="75"/>
  </PersonDescriptor>
  <ObjectDescriptor>
    <TextualDescriptor>Coffe Cup</TextualDescriptor>
    <PositionDescriptor x="50" y="60"/>
  </ObjectDescriptor>
</Image>
</ObjectDescriptor>

```

```

<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet
version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xsl:template match="/">
    <html>
      <head/>
      <body>
        <span style="font-weight:bold">
          Image ID:
        </span>
        <xsl:for-each select="Image">
          <span style="font-weight:bold">
            <xsl:for-each select="@ID">
              <span style="font-weight:bold">
                <xsl:value-of select="."/>
              </span>
            </xsl:for-each>
          </span>
        </xsl:for-each>
      </body>
    </html>
  </template>
</xsl:stylesheet>

```

XSLT



MPEG-7 is the Multimedia Content Description Interface

- ❑ Defined by the Moving Picture Experts Group (MPEG) existing as standardization body since 1988
- ❑ **MPEG-1/MPEG-2** defines audio/video coding, streaming multiplex syntax, and transmission
- ❑ **MPEG-4** extends MPEG-1/MPEG-2 by including graphics elements/higher compression rates/mobile delivery and started off in 1993
- ❑ 1997 started MPEG to work officially on **MPEG-7**

Overview

- ❑ ISO/IEC / Motion Picture Experts Group
- ❑ Multimedia Content Descriptor Interface
- ❑ Opened Standard Based Framework
- ❑ Exchange of Metadata between several parties
- ❑ Currently mainly used in database queries
- ❑ No/small licensing fees like e.g. in MPEG-4
- ❑ Binary + Textual Format
- ❑ Multiplexing in MPEG-2 TS standardized
- ❑ “Metadata Extractors” will be available in the next years
- ❑ This metadata is available for “free”

Scenarios

- Broadcast media selection (push model)
 - Agent based media selection and filtering
- Information access for people with special needs
- Personalization
- Intelligent multimedia presentation
- Personalizeable browsing, filtering, and search for consumers
- Indexing and retrieval of audiovisual archives (pull model)
- Areas
 - Journalism, Entertainment, Education, Surveillance, Remote Sensing, Telemedicine, Bio-Medical Applications, etc.

Standardization Parts of MPEG-7

ISO/IEC 15938

- ❑ **Part 1:** Systems standardizes binary transmission, synchronization and storage modes, ISO/IEC:15938:1:2001
- ❑ **Part 2:** the metadata language is defined in part 2 “Description Definition Language”, ISO/IEC:15938:2:2001
- ❑ **Part 3/4:** visual and audio description schemes are defined in part 3 “Visual” and part 4 “Audio”, ISO/IEC:15938:3:2001, ISO/IEC:15938:4:2001
- ❑ **Part 5:** non-A/V content descriptors are described in part 5 “Generic Entities and Multimedia Description Schemes (MDS)”, ISO/IEC:15938:5:2001
- ❑ **Part 6:** “Reference Software” (ISO/IEC:15938:6:2001) defines general software that supports the different standardized parts
- ❑ **Part 7:** “Conformance Testing” focuses on processes for testing MPEG-7 conformant hardware or software implementations
- ❑ **Part 8:** “Extraction and Use of MPEG-7 Descriptions” defines the multimedia content description interface and the procedures for the use of MPEG-7 tools and the implementation of the reference software

Basic Application Areas

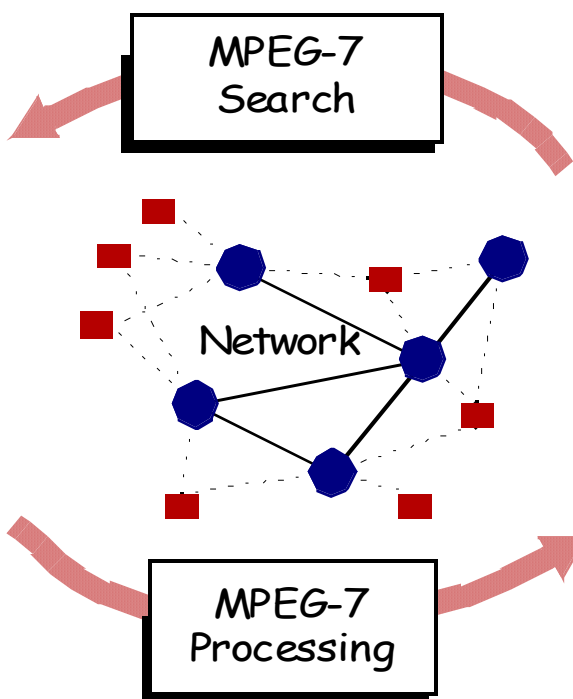
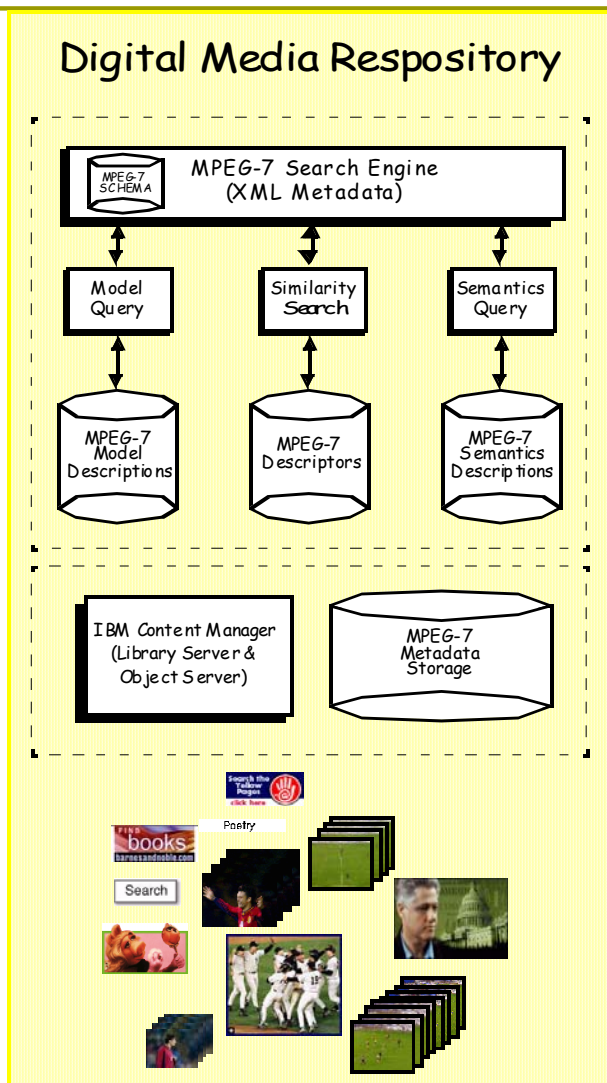
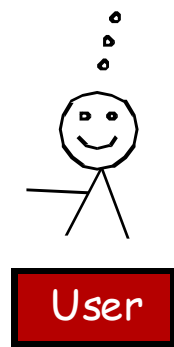
- ❑ Annotation of multimedia content for search, retrieval and management of content
 - ❑ MPEG-7 can be characterized as a catalyst and type set source for newly emerging standards
 - ❑ System implementation for textual representation and binary streaming of metadata
- ❑ The MPEG-7 key-concepts
 - Format independence
 - Minimal standardized part definitions
 - Development of techniques for extraction, encoding and use is left to industry
 - Challenge of automated metadata extraction

MPEG-7 Multimedia Indexing and Searching

□ MPEG-7 Indexing & Searching:

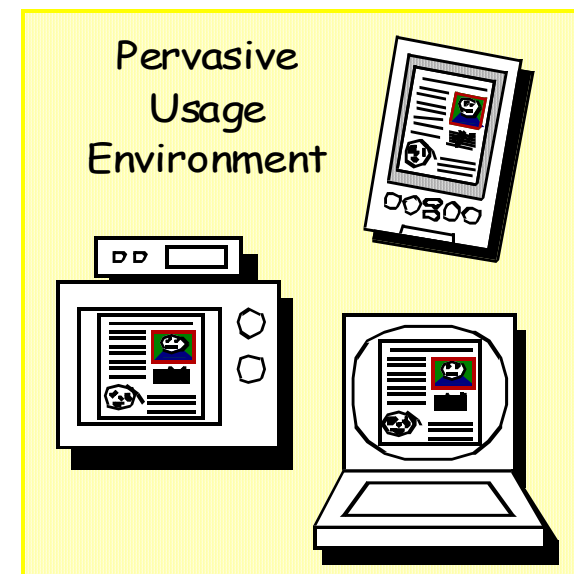
- Semantics-based (people, places, events, objects, scenes)
- Content-based (color, texture, motion, melody, timbre)
- Metadata (title, author, dates)

Sounds like ...
Looks like ...



□ MPEG-7 Access & Delivery:

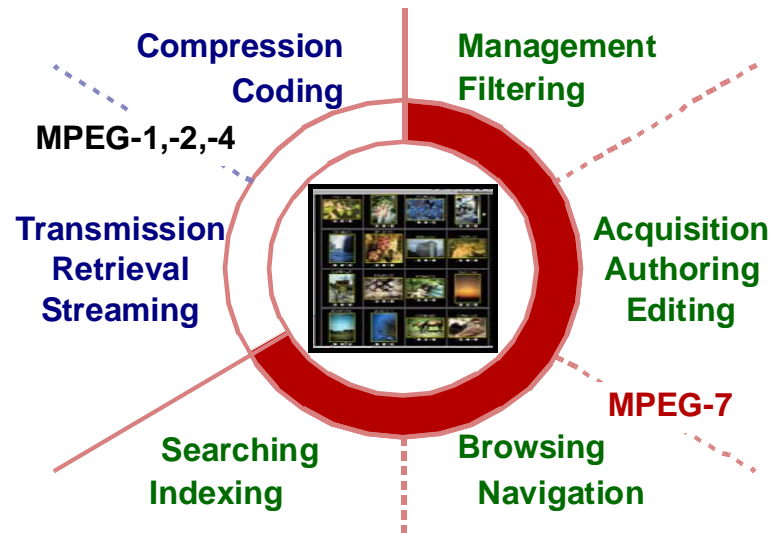
- Media personalization
- Adaptation & summarization
- Usage environment (user preferences, devices, context)



Source: MPEG-7 Multimedia Content Description Standard, John R. Smith, IBM, Presentation online, January 8, 2003

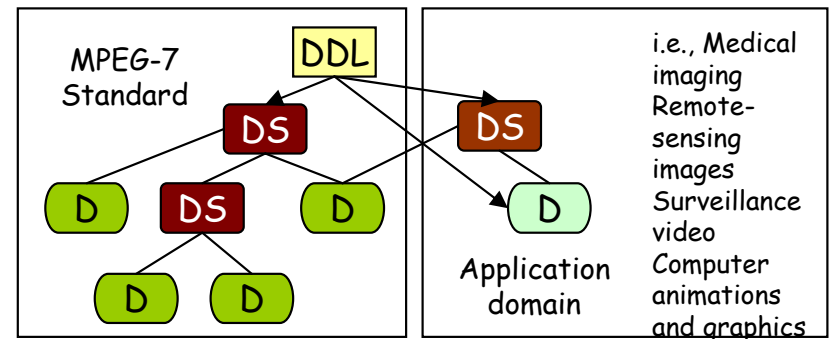
MPEG-7 Overview

(XML for Multimedia Content Description)



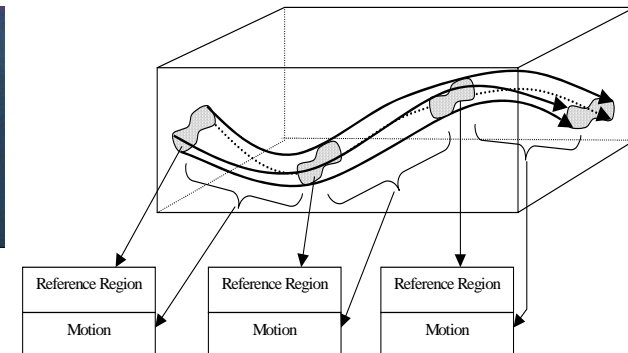
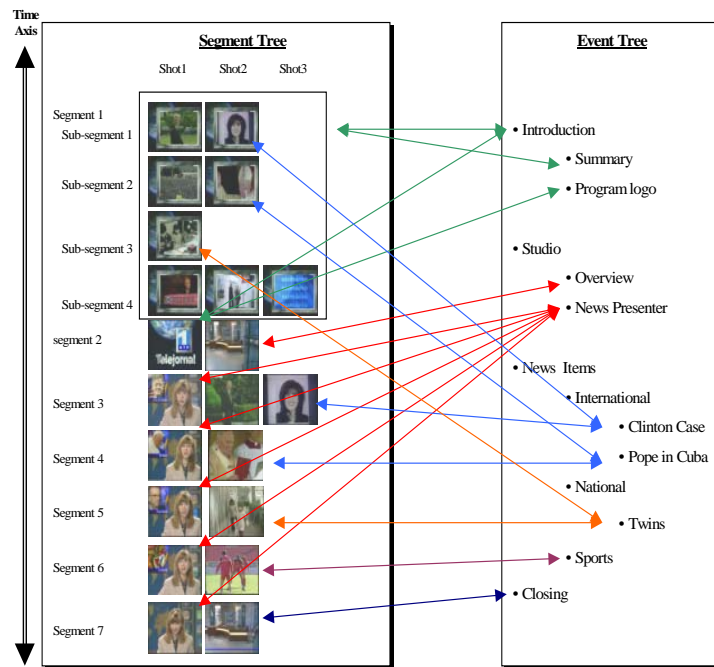
□ MPEG-7 Normative elements:

- Descriptors and Description Schemes
- DDL for defining Description Schemes
- Extensible for application domains



□ Rich, highly granular multimedia content description:

- Video segments, moving regions, shots, frames, ...
- Audio-visual features: color, texture, shape, ...
- Semantics: people, events, objects, scenes, ...

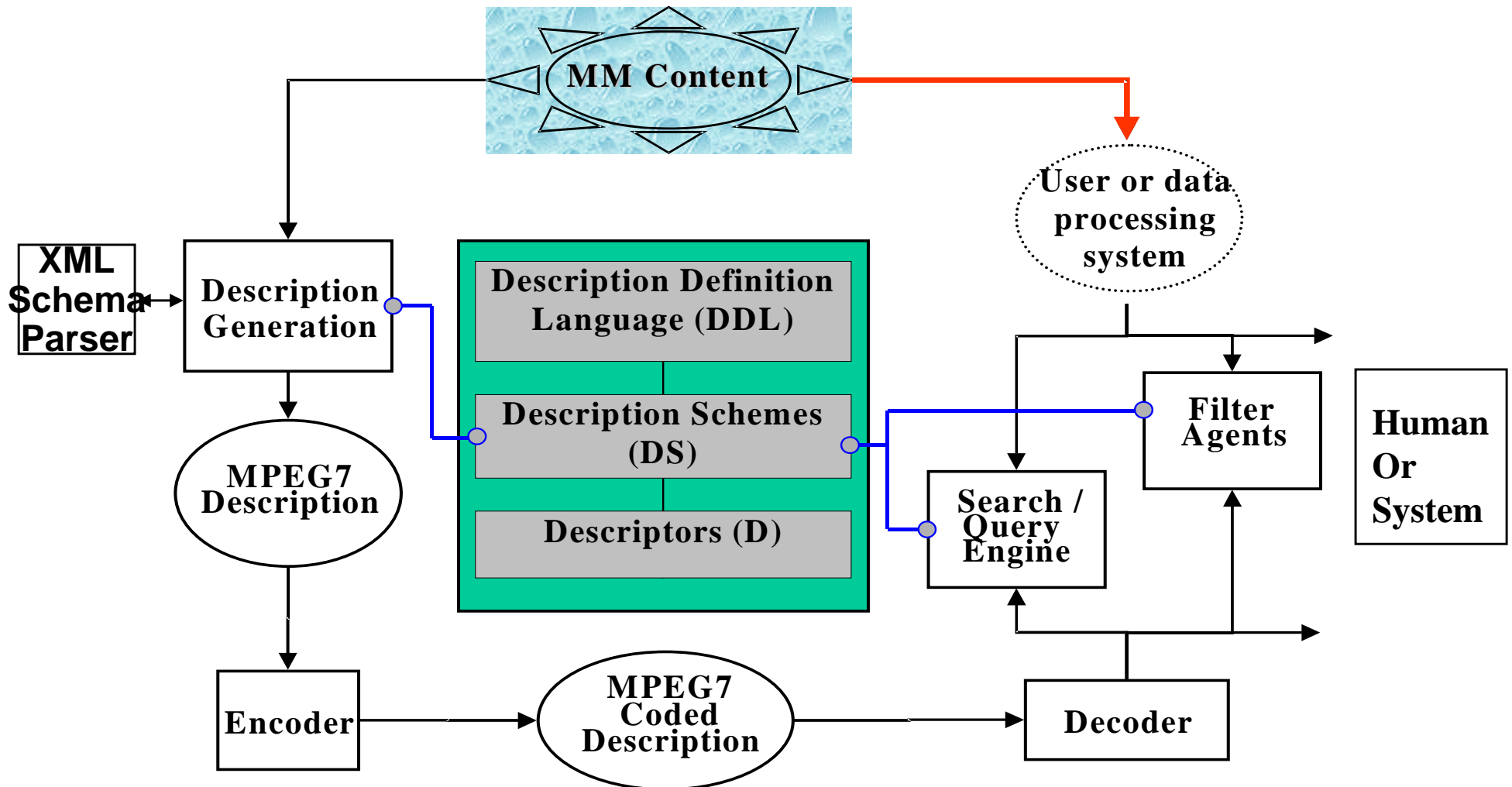


MPEG-7 Metadata Definitions

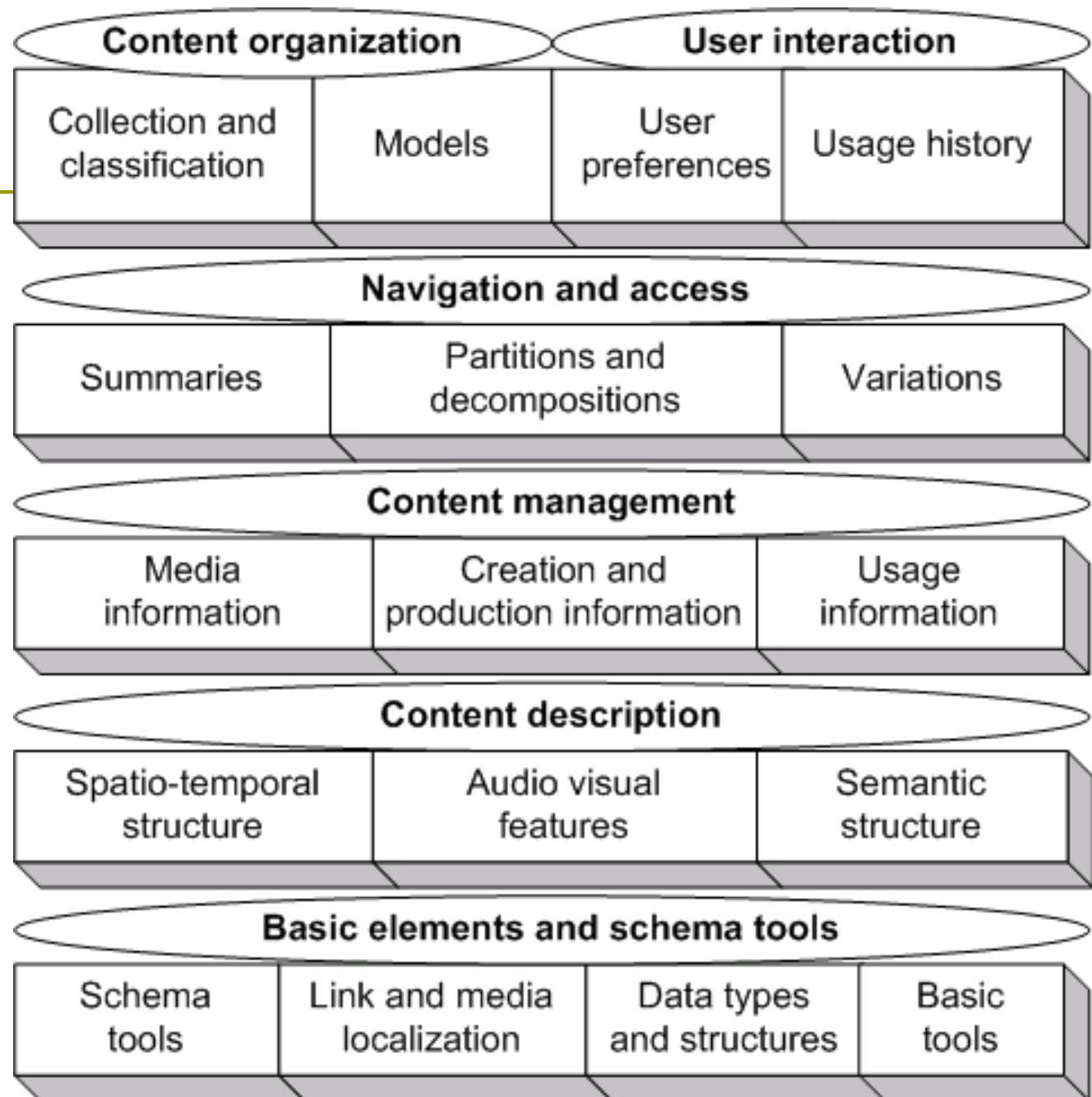
- **Descriptor Definition Language (DDL):**
 - XML Schema with slight extensions
- **Multimedia Description Language (MDL):**
 - General abstract features
- **Visual and Audio Metadata Definitions**
 - Features of audio/visual content

- *Descriptor*: syntactic/semantic representation of features (e.g. colour)
- *Definition Scheme*: rules, relations and semantics between either descriptors/definition schemes (e.g. spatio-temporal structure)

Systems



MPEG-7 Metadata Definitions



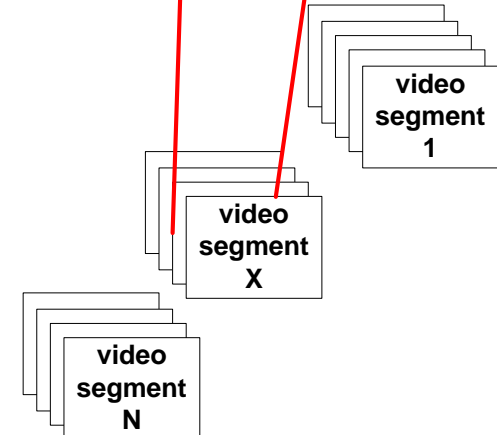
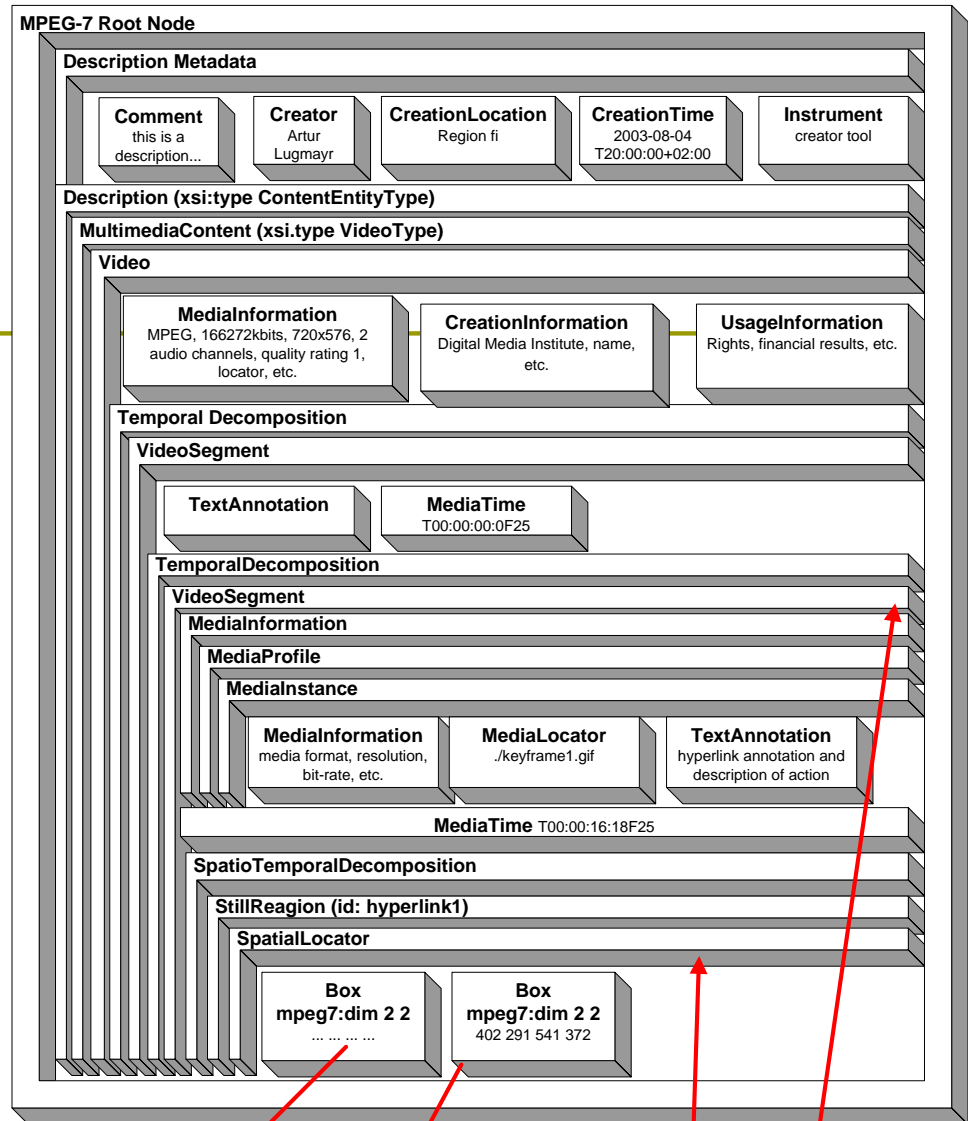
Basic Elements and Schema Tools

- ❑ Describe general valid metadata definition
- ❑ On very abstract level
- ❑ Several parts of MPEG-7 standards use these descriptors as base to build newly defined descriptors
- ❑ Consist of:
 - BASIC TOOLS: complex and more application specific metadata definitions (e.g. textual descriptors)
 - SCHEMA TOOLS: describe the process for creating valid MPEG-7 metadata instances
 - DATA TYPES/STRUCTURES: describe general layout of MPEG-7 documents (e.g. probability vectors)
 - LINKING/LOCALIZATION: implicit/explicit identification reference and localization methods (e.g. media time)

Annotating Multimedia Content

- ❑ Structural decomposition of multimedia assets
- ❑ Spatio-temporal/hierarchical decomposition of multimedia assets
- ❑ Include semantic/narrative meaning
- ❑ How?
 - Events (e.g. a happening) occur to objects (e.g. people) and are triggered according to abstract conceptual models (e.g. state diagram)

Example



Example

MPEG-7 Root Node

Description Metadata

Comment, Creator, Creator Location, Creation Time, Instrument

Description

Multimedia Content

Video

Media Information, Creation Information, Usage Information

Temporal Decomposition

Grouping Multimedia Assets: Content Organization

- ❑ Group or cluster multimedia assets by certain criteria or collections
- ❑ Criteria:
 - Logical packaging (e.g. folder structure)
 - Higher complexity (e.g. movie genre)
- ❑ Toolsets:
 - *Collections*: describe groups/collections, their subgroups and relations
 - *Models*: associate characteristics with instantiated groups of multimedia
- ❑ Example: Subjective Image Quality Estimation
 - Image quality determination with histogram distribution
 - Low quality = less clustered histogram
 - High quality = equally clustered histogram

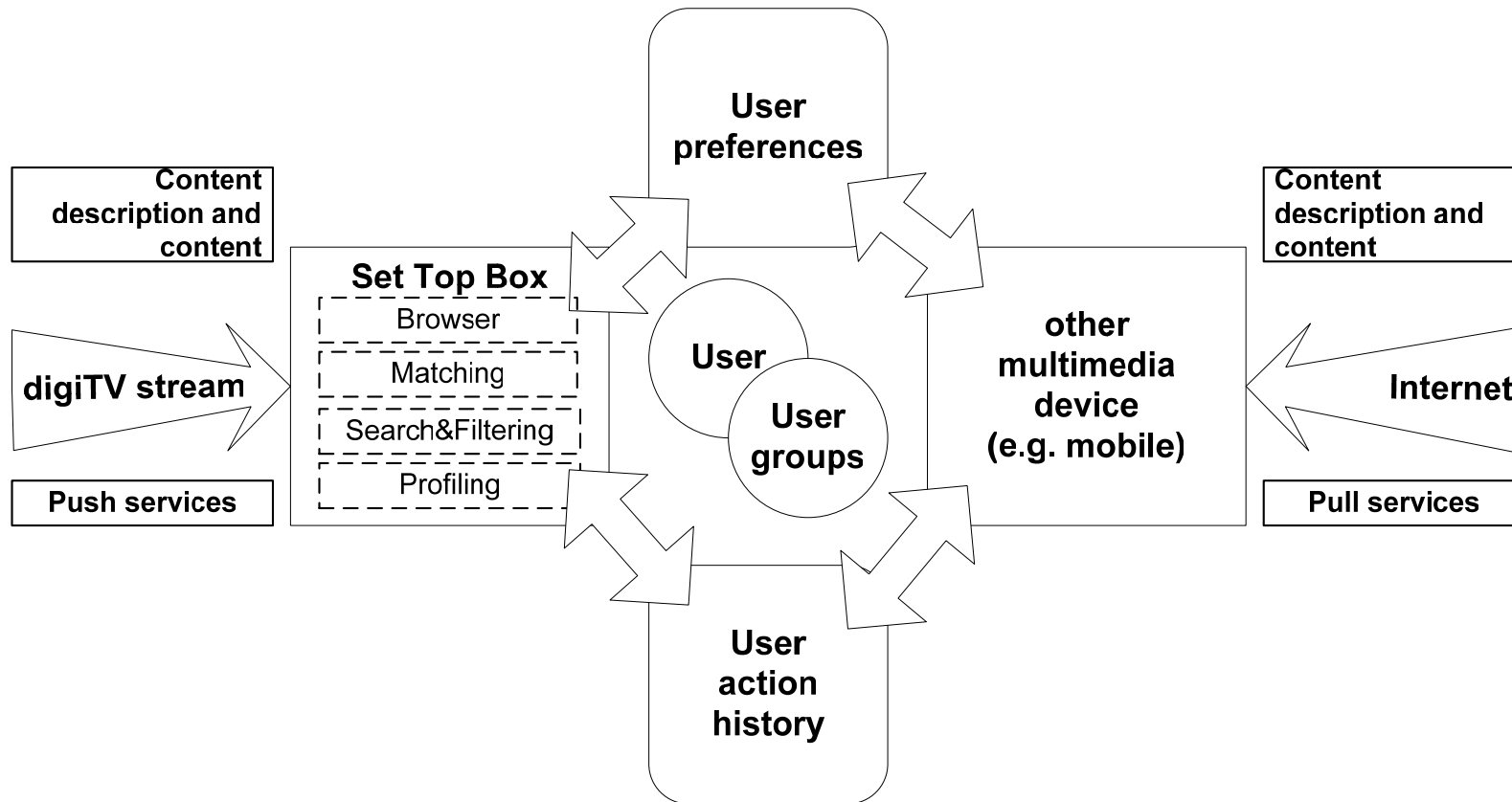
Managing Conventional Media Archive Information

- Content management tools to describe the whole life-cycle of multimedia assets
- Types:
 - *Media information*: independent identification of unique multimedia assets (e.g. coding format)
 - *Creation and production information*: hold data about the overall process of creating multimedia assets (e.g. genre, actors)
 - *Content usage information*: describe rights of usage (e.g. financial information)

Easy Navigation & Access

- ❑ Modes for accessing and browsing through multimedia databases
- ❑ Enable features, summaries, partitions, decompositions, and variations
- ❑ **Summaries** = hierarchical decomposition of multimedia content for preview
- ❑ **Partitions** = spatial-temporal inner media navigations
- ❑ Adaptation of existing material to local facilities and user browsing habits

Personalization, User Interaction and Consumer Profiles



Audio Descriptors

- ❑ Tools for audio descriptions
- ❑ Similarity features: timbre, melody, harmony...
- ❑ For music, speech, instruments, spoken content

Table 3.5. MPEG-7 audio descriptors [114]

MPEG-7 Descriptor	Features
low-level descriptors	temporal and spectral descriptors grouped into basic, basic spectral, signal parameters, temporal timbral, spectral timbral and spectral basic descriptors
high-level descriptors	descriptors for sound recognition and indexing, spoken content descriptors, melody descriptors and musical instrument timbre descriptions
spoken content descriptors	speaker information, spoken content header, word and phone lexicon, spoken content lattice
sound classification and similarity	spectral basis functions, sound classification schemes, sound probability models, sound model histogram, sound model state path

Visual Descriptors

- Describe any type of visual content
- Similarity measures are also described
- Video, images, 3D, 2D, etc.

Table 3.6. MPEG-7 visual descriptors [118]

MPEG-7 Descriptor	Features
color descriptors	color spaces (e.g. HSV), dominant color, scalable color, color structure, color layout, group-of-frames color, group-of-pictures color
texture descriptors	homogeneous texture, texture browsing, edge histogram
shape descriptors	region-based descriptors, contour-based descriptors, 3D shape
motion descriptors	motion trajectory, motion characterization, motion activity, camera motion, parametric motion

MPEG-7 Systems

□ Defines:

- Streaming metadata in parallel to audio/video content
- Metadata protocol stack model
- Optimized metadata representation

□ Features:

- Abstract MPEG-7 system architecture
- Optimized binary format for metadata
- Support for any transmission mode (push/pull)
- Bi-directional mapping of metadata

MPEG-7 Systems – Why binary?

- ❑ Binary metadata enables synchronization of metadata trees between different devices
- ❑ Partial or complete metadata tree updates
- ❑ Optional data encryption for increased security
- ❑ Validation checks for well-formed-ness
- ❑ Can be used for any XML based metadata

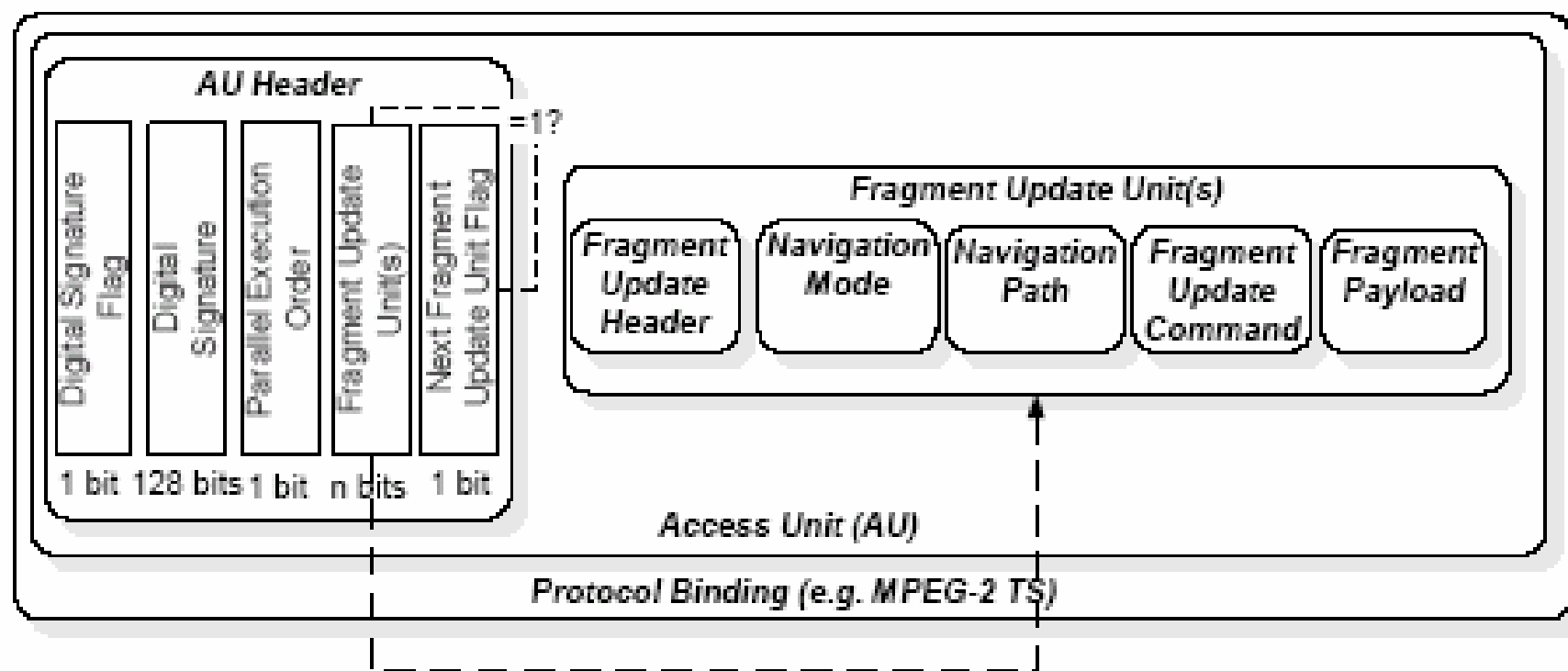
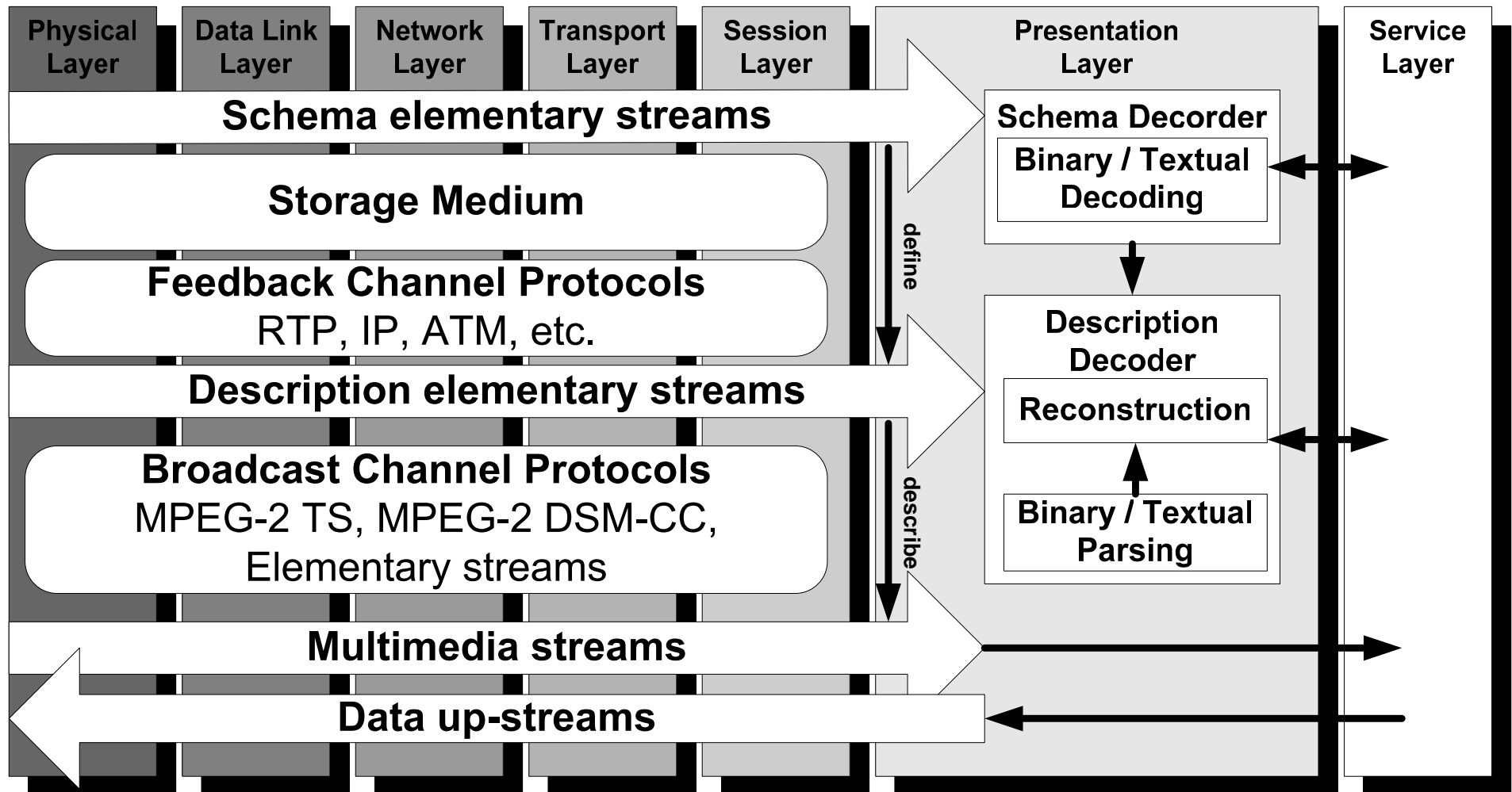


Fig. 5.9. BiM data packet consisting of *access units (AUs)*, *fragment update units (FUUs)* as payload and the carrying protocol (e.g. MPEG-2 TS) based on MPEG-7 standards

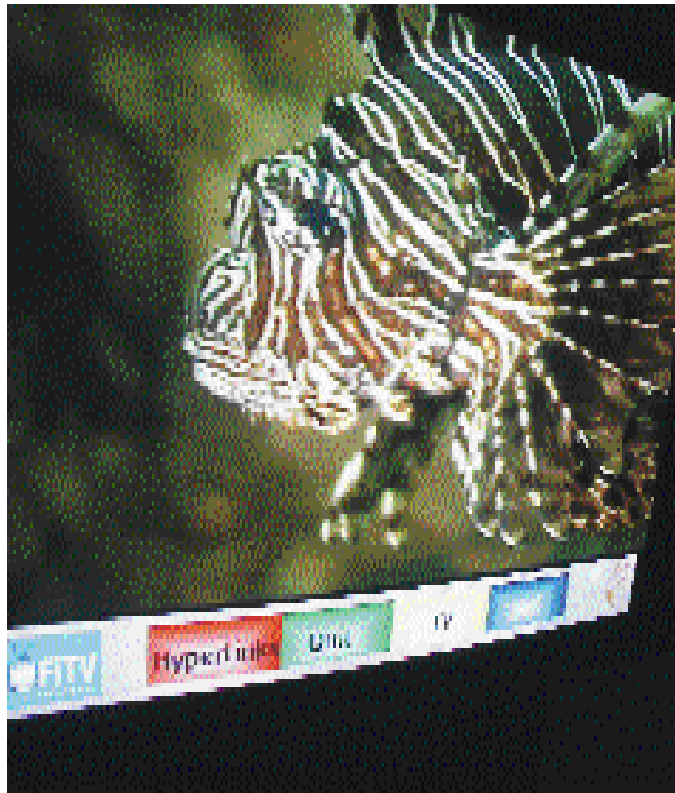
MPEG-7 Systems - Functionality

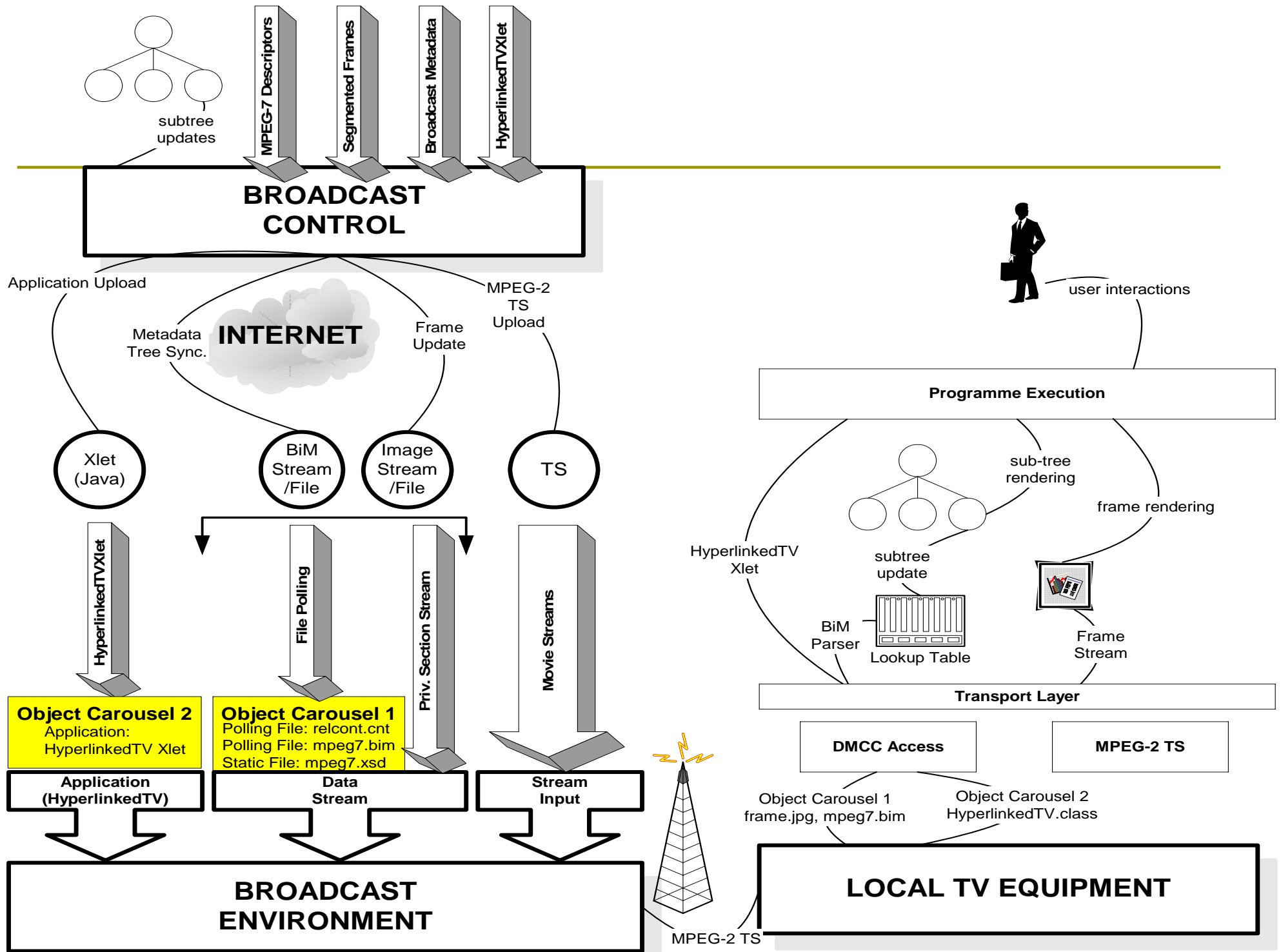
- ❑ Mapping complete or partial metadata trees into so-called Access Units (AUs)
- ❑ Each AU consists of header (equal mechanism to IP packets on the Internet)
- ❑ Each AU carries one or more fragment update units (FUUs)
- ❑ FUU contains information about navigation path, insertion modulus and how the update has to be performed

Metadata Protocol Stack



USE CASE: Hyperlinked TV





Special deployment issues - Synchronization

	Entity of Operation	Example	Characteristics
Media Layer	Single continuous metadata stream	PES BiM Application Stream	Device independent interface of operations en-/decapsulation processes
Stream Layer	Group of media streams Single Media Stream	MPEG-2 DVB stream A whole program inc. applications and content	Interstream synchronization Intrastream synchronization
Object Layer	Temporal synchron. Specification as input	Multimedia presentation Hyperlinked TV including object information	Intra-/Inter stream synchronization based on lower layer calls
Transaction Layer	Higher layer Internet protocols (HTTP, etc.)	XML based eBusiness	Asynchronous / Synchronous
Human Layer	Human interactions	Activation of Hyperlinked TV	Asynchronous
Material Preparation Layer	Preoperational entities	Extraction of metadata	Soft deadlines

USE CASE: System Configuration

Functionality	Technology	Comments
Service Front End	Apache HTTP Server V1.3.23	Web-server
	Apache Tomcat V4.0.2	Servlet container deployment applications
	Apache Struts V1.0.2	Tag library for providing a servlet deployment environment
Service Control	Servlet technology	Library for servlet implementations for different services, and a generic general deployment framework
Multimedia Asset Repository	Tamino XML Server	Metadata native storage system
	MySQL V3.23.52	Open source relational database system
Communication Protocols	SOAP V2.3.1	Communication in XML between front-end and consumer
	XML Binarizer V0.9 (own development of a MPEG-7 Systems subset)	own development of a MPEG-7 system subset
	UDP, TCP, IP, HTTP, RTP	Standard deployment protocols for web-based services; RTP is used for video streaming, others depend on the deployed service
XML Editing Environment	XMLSpy IDE	

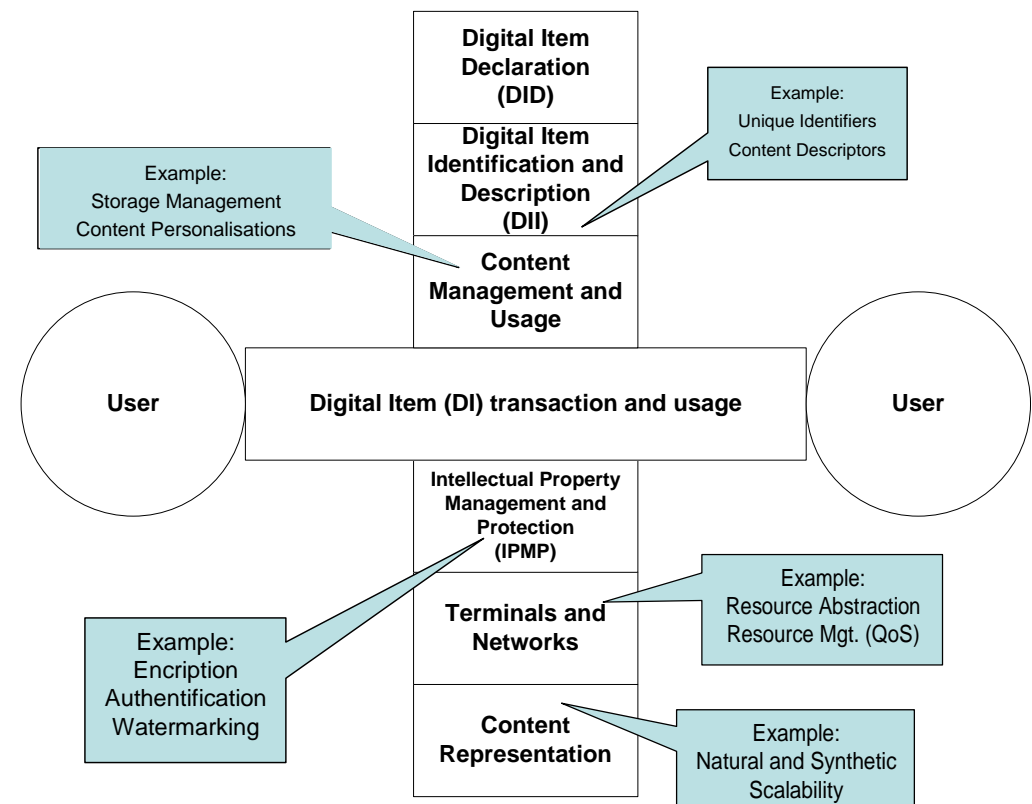
The steps after MPEG-7...

- MPEG-21...
- Transition from Ambient Multimedia to Bio-Multimedia...

MPEG-21 — A MULTIMEDIA FRAMEWORK



- ❑ 'Umbrella' multimedia framework throughout the value-chain
- ❑ Work started in June 2000
- ❑ ISO/IEC JTC1/SC29/WG11 MPEG
- ❑ Complementary tools for MPEG-1/2, MPEG-4 and MPEG-7
- ❑ Worldwide standard approx. by the end of 2005
- ❑ Valid throughout the value-chain – from creation to consumption
- ❑ 'First-mover' provide already tools & applications

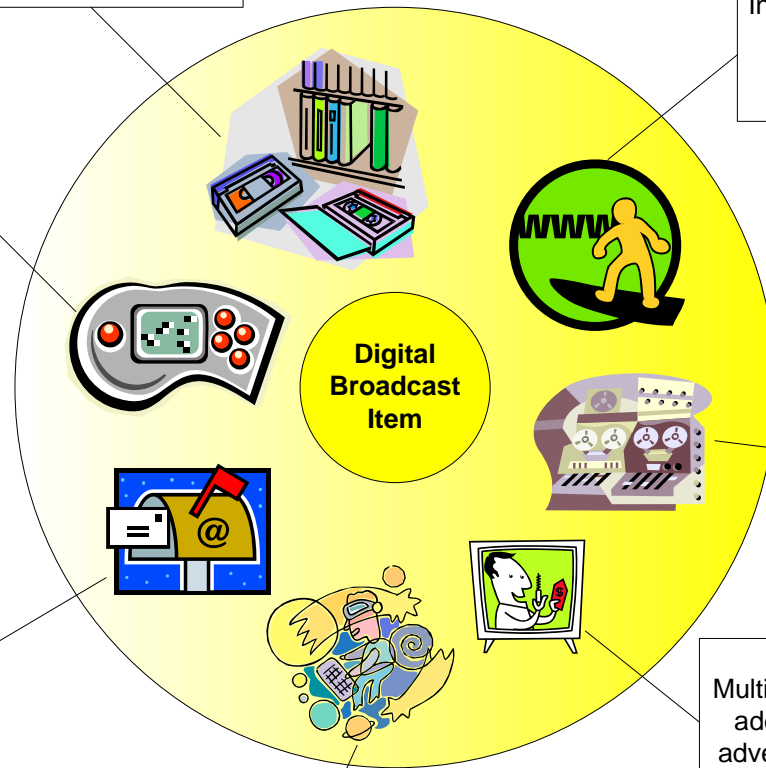


ASSETS = Digital Item

Audio/Video Content in different profiles (e.g. Dolby Surround, Stereo, Mono, High / Low-resolution)

Internet and Web content (DVB-HTML, HTML, XML)

Software and Applications (e.g. computer games, set-top-box updates, chatting software)



Technical data (e.g. synchronization information, real-time information)

Communication facilities (e.g. email, chat)

Multimedia asset add-ons (e.g. advertisements)

Interactivity facilities (e.g. PDA, remote control, mobile)

- ❑ Adaptation
- ❑ Digital Rights Management
- ❑ Rights Expression
- ❑ Media Management
- ❑ Packaging
- ❑ Processing
- ❑ Communication

Research Challenges of AmI

Components

Ambient

- Smart materials
- Micro/Nano-electronics/systems
- Smart materials
- MEMS technology & sensor technology
- Embedded systems
- Ubiquitous Communication
- I/O device technology
- Adaptive software

Intelligence

- Media management & handling
- Natural interaction
- Computational intelligence
- Contextual awareness
- Emotional computing

Platform design

Software design, engineering and integration
Experience prototyping

Questions???

lartur@acm.org

Slides can be found on:
www.digitalbroadcastitem.tv