Introduction MPEG-7 by T-Nova Berkom

MPEG-7 – the new international Standard Multimedia Content Description Interface

author: Ingo Wolf

contact: wolfi@t-systems.com

===!"§==Systems=

Agenda

Session 1

- motivation: MPEG standards, why MPEG-7 is needed
- scope and objectives
- fields of application
- structure of the standard – part 1-2 of MPEG-7

- - - - - - - - B r e a k - - - - - - - -

Session 2

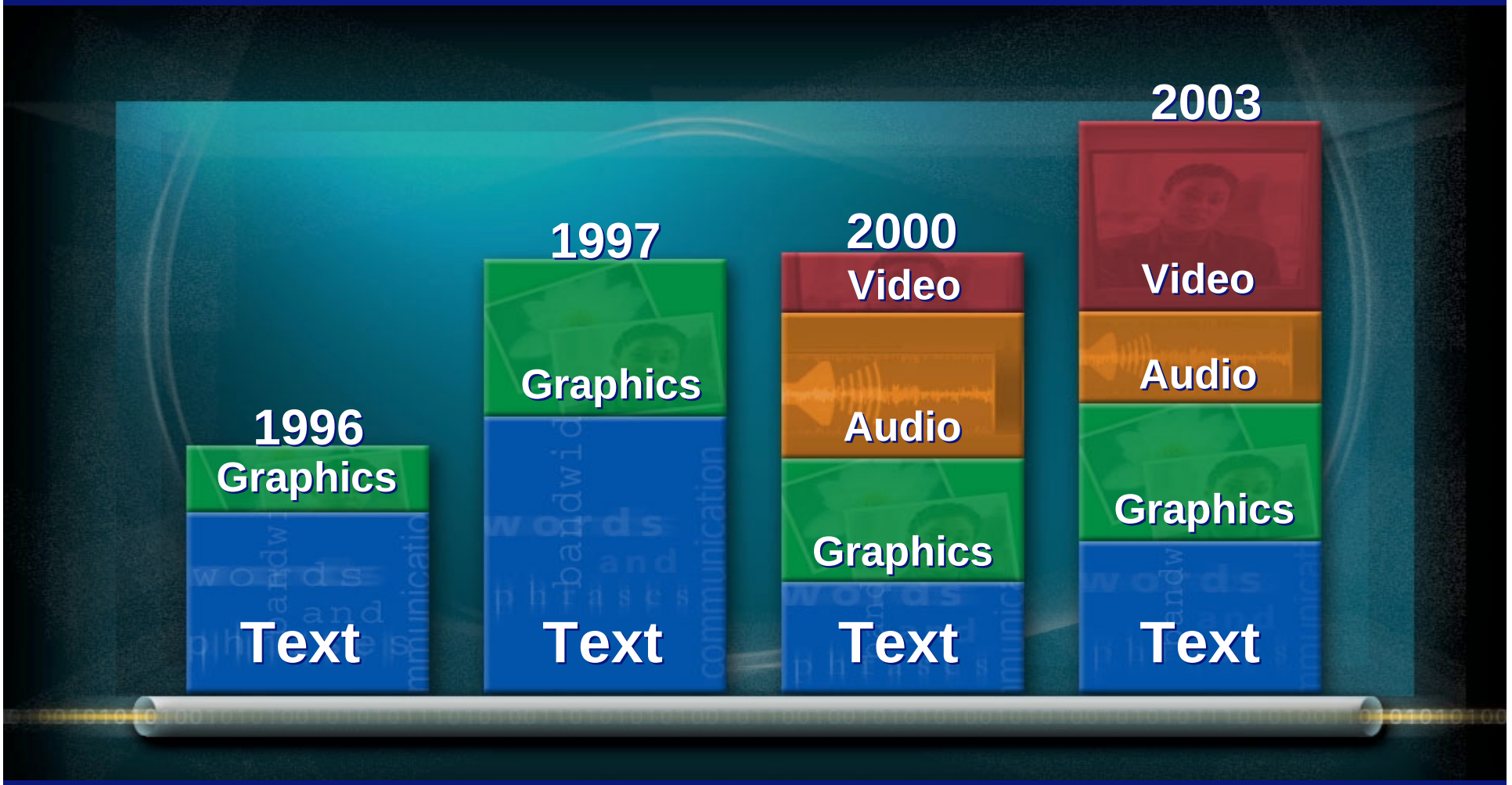
- structure of the standard –parts 3-8 of MPEG-7
- links and hints on MPEG-7 activity

Motivation – MPEG standards

ISO/IEC SC29Wg11 Standards Group began in 1988. Today, over 300 developers attend working conferences (average 4 conf. per year) - over 35 national delegations vote.

- ***MPEG-1:*** For the storage and retrieval of moving pictures and associated audio on storage media. (11/1992)
- ***MPEG-2:*** For digital television, it's the timely response for the satellite broadcasting and cable television industries in their transition from analog to digital formats. (11/1994)
- ***MPEG-4:*** Codes content as objects and enables those objects to be manipulated individually or collectively on an audiovisual scene. (vers.1: 09/1998 vers.2: 11/1999)
- MPEG-1, -2, and -4 make content available. MPEG-7 lets you to find the content you need. (09/2001)

Motivation - why MPEG-7 is needed



Data: Inktomi, USA

Motivation – explosion of multimedia content requires interoperable applications

growth of multimedia content

soap operas



sport



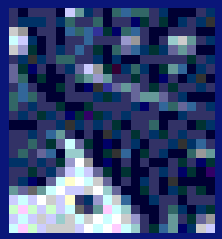
news



private content



scientific content



requires

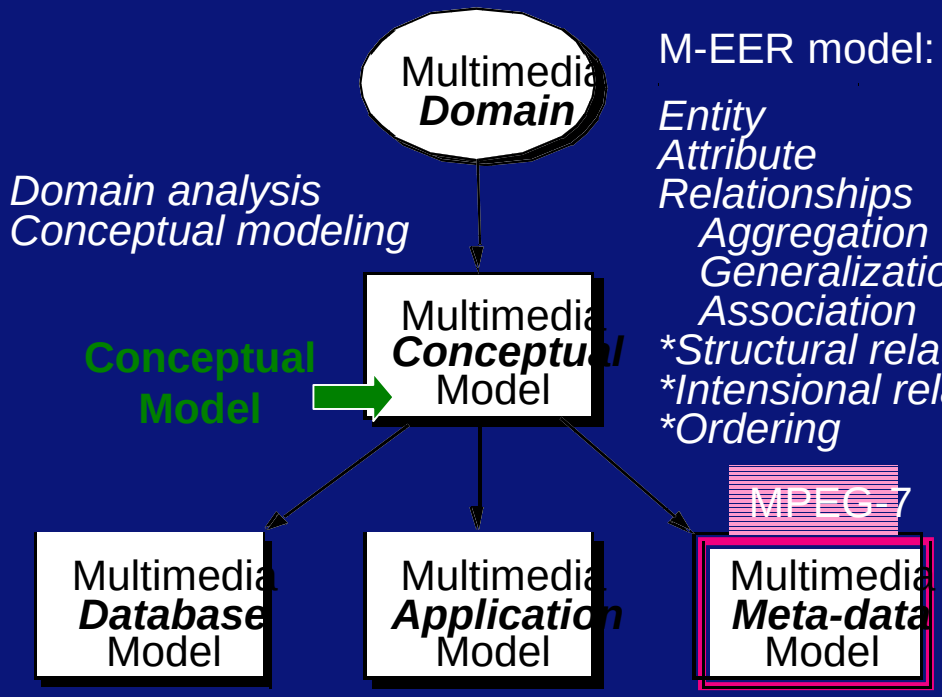
efficient, intelligent, interoperable applications

- storage and retrieval
- universal multimedia access
- broadcast media selection
- personalized TV
- E-Commerce and Tele-shop
- games and galleries
- B2B-Content-Management
- multimedia editing
- surveillance and security applications
- Etc

===!"§==Systems=

Scope and objectives of MPEG-7

Conceptual Modeling for Multimedia Content Description



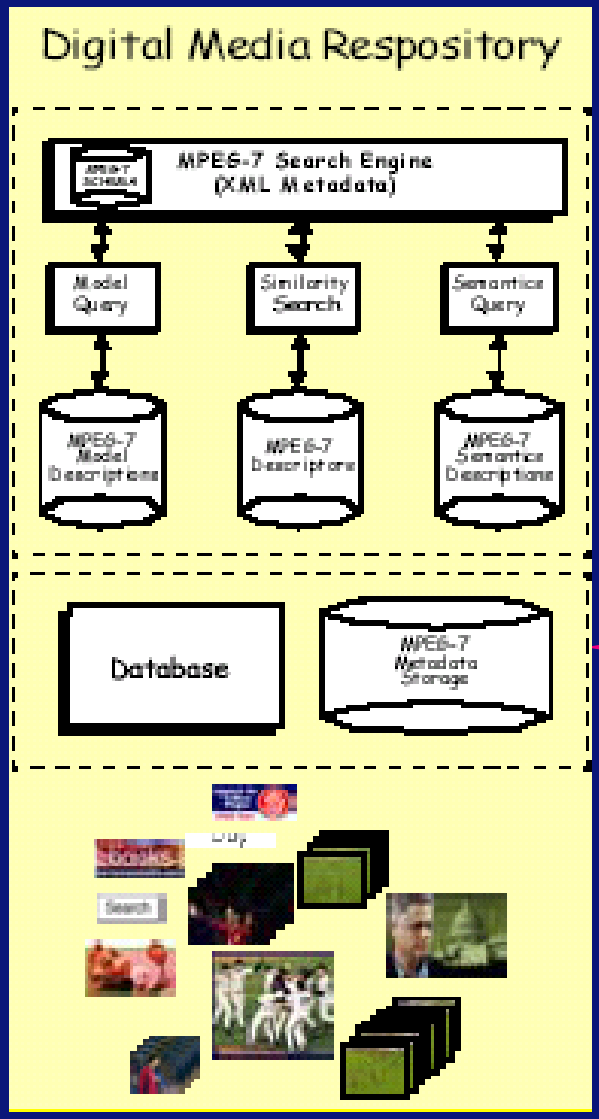
Modeling: objective of creating single unified model (conceptual, logical, implementation)

Concepts: multimedia domain analysis

Transformation: of conceptual model into the logical models for

- Database design
- Software classes for application implementation
- Meta-data language for interoperable annotation

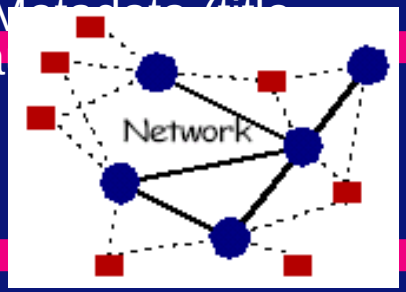
Scope and objectives of MPEG-7



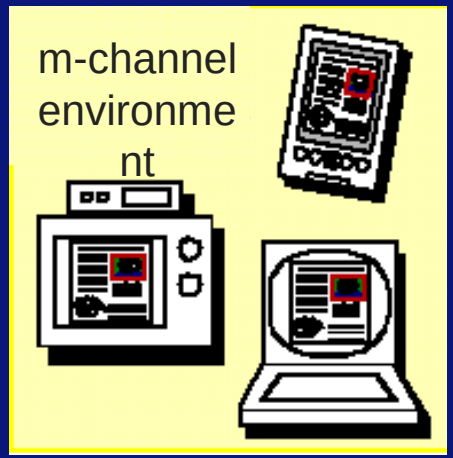
- **MPEG-7 Indexing & Searching:**
  - Semantics-based (people, places, events, objects, scenes)
  - Content-based (color, texture, motion, melody, timbre)
  - Metadata (title, a







- **MPEG-7 Access & Delivery:**
  - Media personalization
  - Adaptation & summarization
  - Usage environment (user preferences, devices,

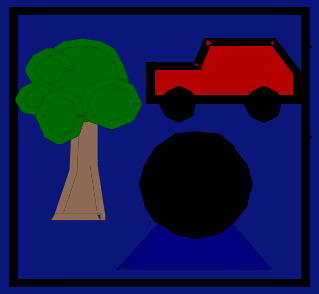


===!"§==Systems=

Scope and objectives of MPEG-7

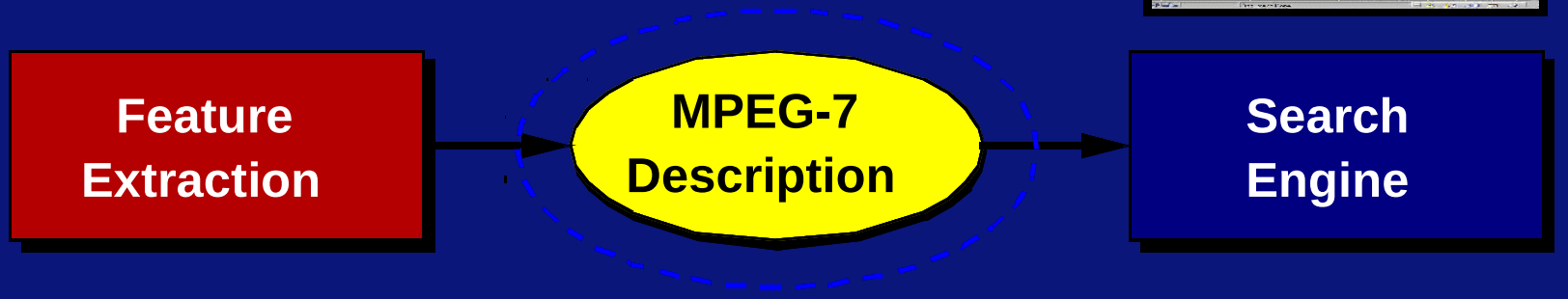
- fast + efficient search, indexing and identification
- separation of content and metadata → independence of description-

  and source-format



standardization

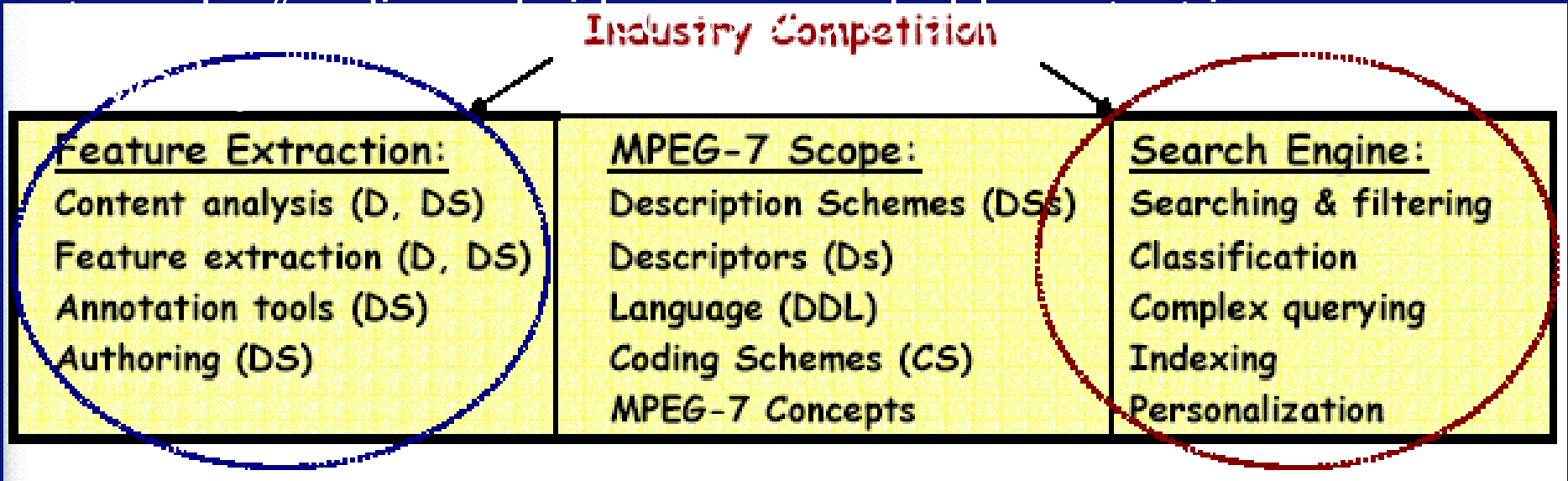




Scope and objectives of MPEG-7

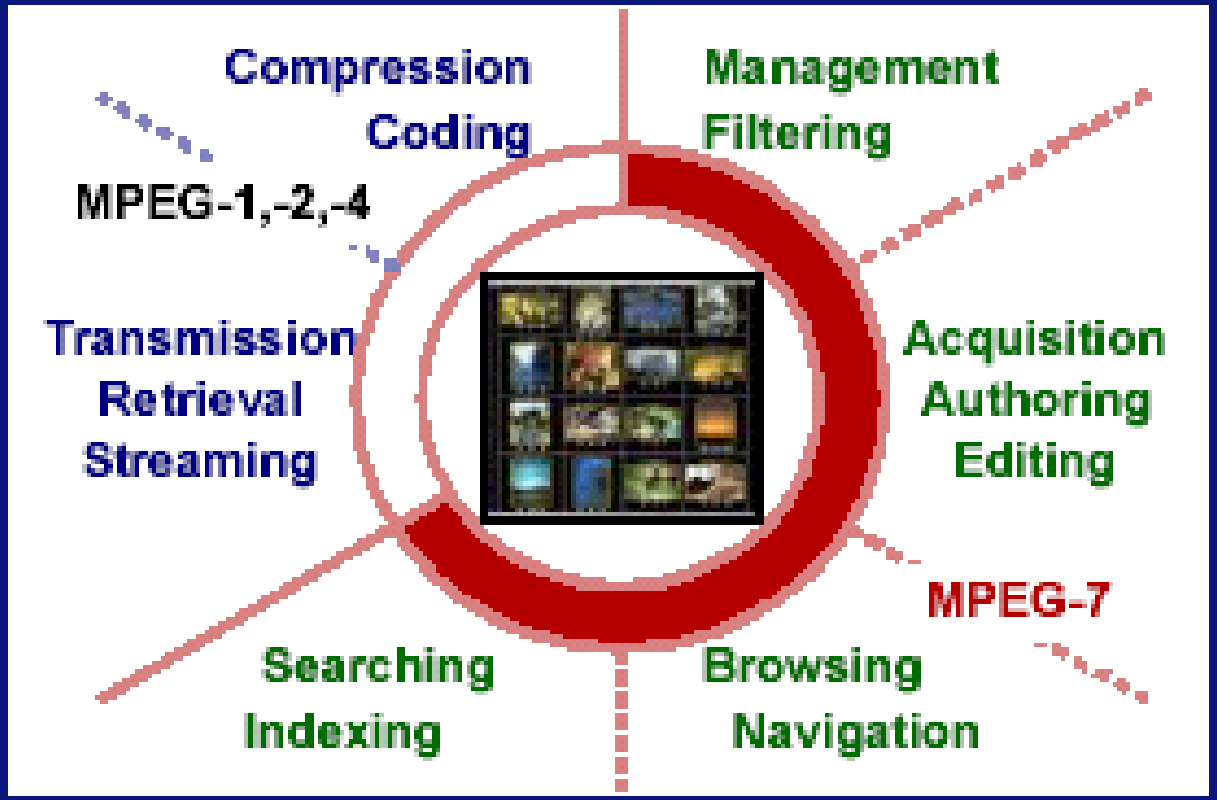
•variety of feature descriptions (low-level features, structure, semantic,

models, collections, creation, etc.) allow a wide range of applications



Fields of application with MPEG-7

XML for Multimedia Content Description



Fields of application with MPEG-7



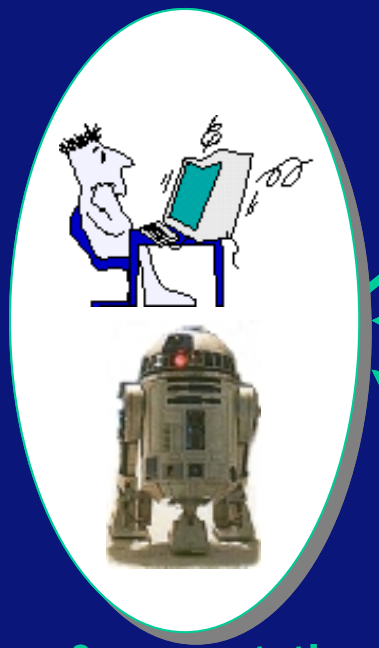
Feature extraction

Manual / automatic

AV Description

Storage

Decoding (for storage)



Search / query

Pull

Browse

Conf. points

Transmission

Encoding (for transmission)

Filter

Push

User & computational systems

Fields of application with MPEG-7

- Architecture, real estate, and interior design (e.g., searching for ideas).
- Broadcast media selection (e.g., radio channel, TV channel).
- Digital libraries (e.g., image catalogue, musical dictionary, bio- medical imaging catalogues, film, video and radio archives).
- E-Commerce (e.g., personalized advertising, on-line catalogues, directories of e-shops).
- Education (e.g., repositories of multimedia courses, multimedia search for support material).
- Home Entertainment (e.g., systems for the management of

Fields of application with MPEG-7

- Journalism (e.g. searching speeches of a certain politician using his name, his voice or his face).
- Multimedia directory services (e.g. yellow pages, Tourist information, Geographical information systems).
- 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).
- Social (e.g. dating services).
- Surveillance (e.g., traffic control, surface transportation, non-

example application with MPEG-7

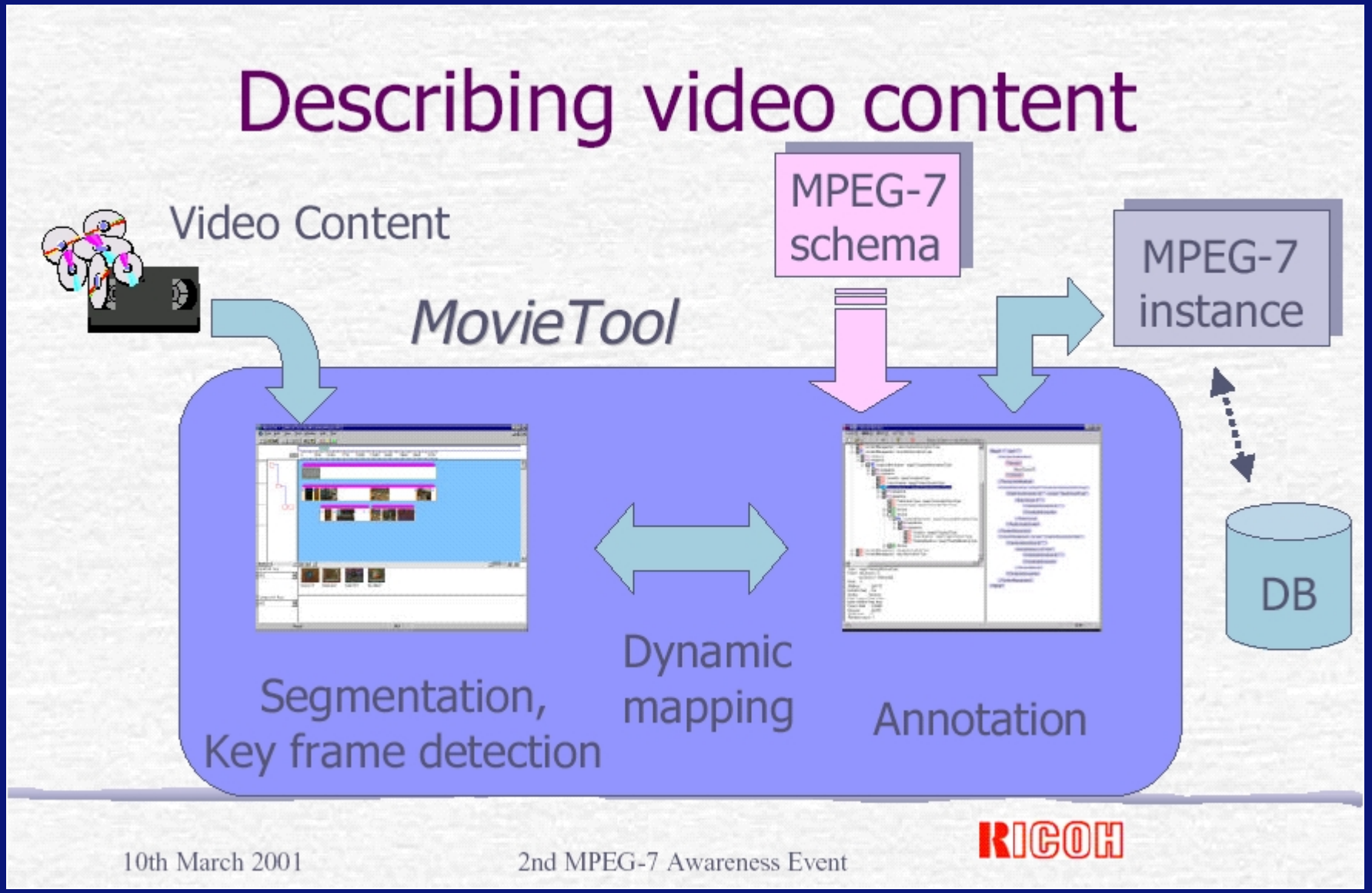
MPEG-7 Camera Application - Smart Fire Alert

(Fastcom Technologies, Switzerland)



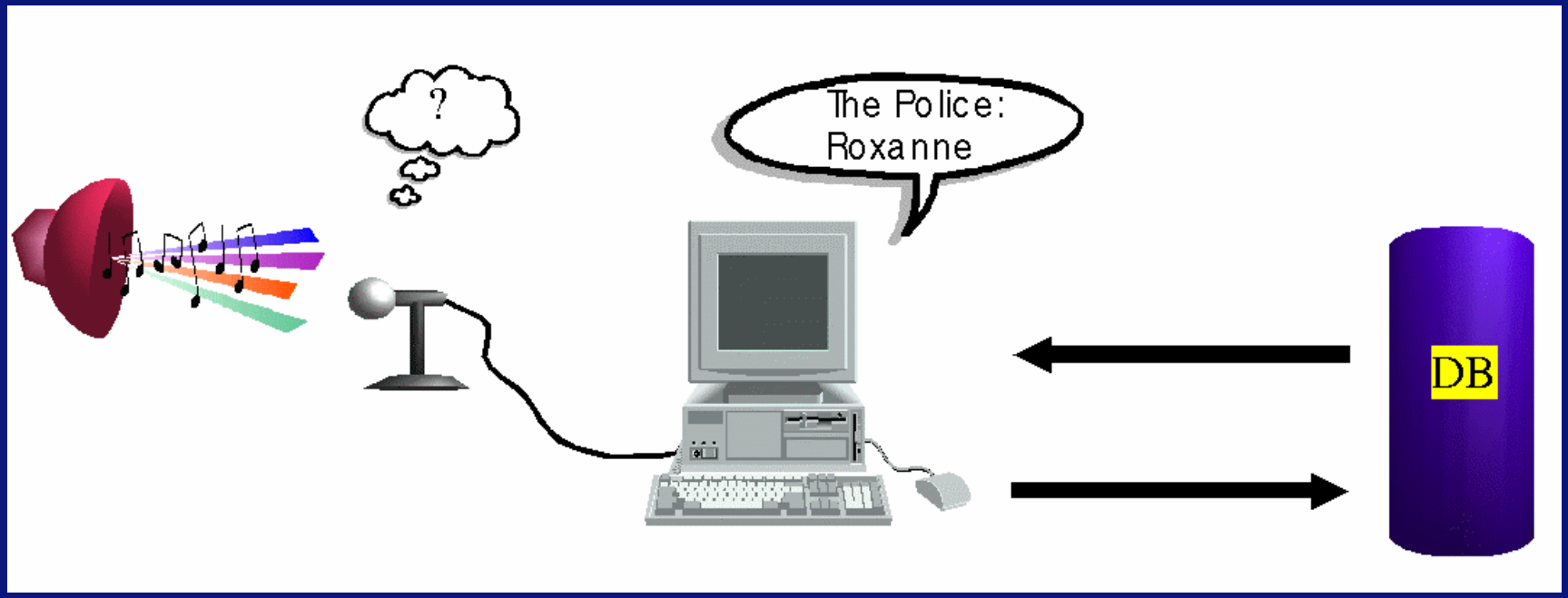
===!"§==Systems=

example application with MPEG-7



example application with MPEG-7

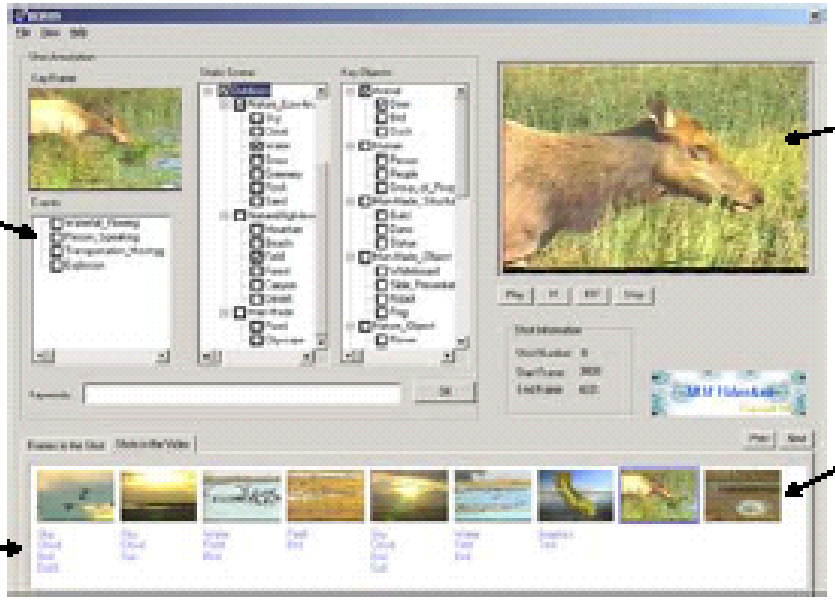
Audio-ID by Frauenhofer-IIS Erlangen



example application with MPEG-7

MPEG-7 Annotation Tool

- MPEG-7 Video Annotation Tool:
  - Semi-automatic tool for annotating video
  - Creates MPEG-7 semantic description metadata
  - Supports customizable semantic classification schemes
  - Exploring active learning framework for propagating annotations
- Available for download from IBM alphaWorks:
  - http://www.alphaworks.ibm.com

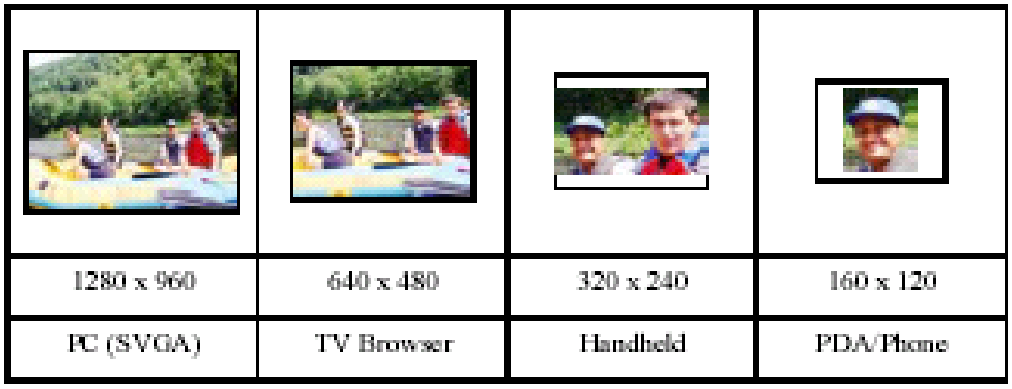


example application with MPEG-7 Mobile Content Adaptation for Images

MPEG-7 annotation image (transcoding hints)



Device display (automatically adapted image)



MPEG-7 region annotation (Ri):

– **Importance Hint (Ii):**
Indicates relative importance of region

– **Spatial Resolution Hint (Si):**
Indicates minimum resolution of region for preserving details

– Selects best combination of cropping and rescaling for device screen Device display

===!"§==Systems=

Structure of MPEG-7: overview parts 1-8

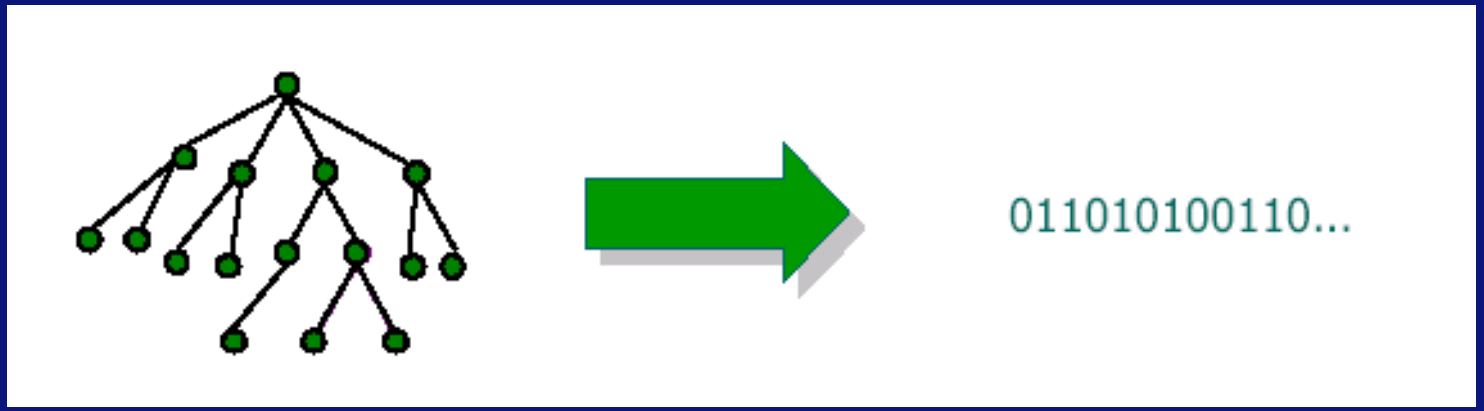
- Part 1: Systems (binary syntax)
- Part 2: Description Definition Language (extension opportunities)
- Part 3: Visual (...)
- Part 4: Audio (...)
- Part 5: Multimedia Description Schemes (core Description Schemes)
- Part 6: Referene Software (C++/Matlab Framework on Linux and

  Windows)
- Part 7: Conformance (conformance and interoperability testing)

==I"§==Systems=

Structure of MPEG-7: part 1 Systems

- specifies the tools for preparing descriptions for efficient transport and storage, compressing descriptions, and allowing synchronization between content and descriptions.
- **Systems tools** support delivery of descriptions, multiplexing of descriptions with multimedia content, synchronization, file format, and so forth.



===!"§==Systems=

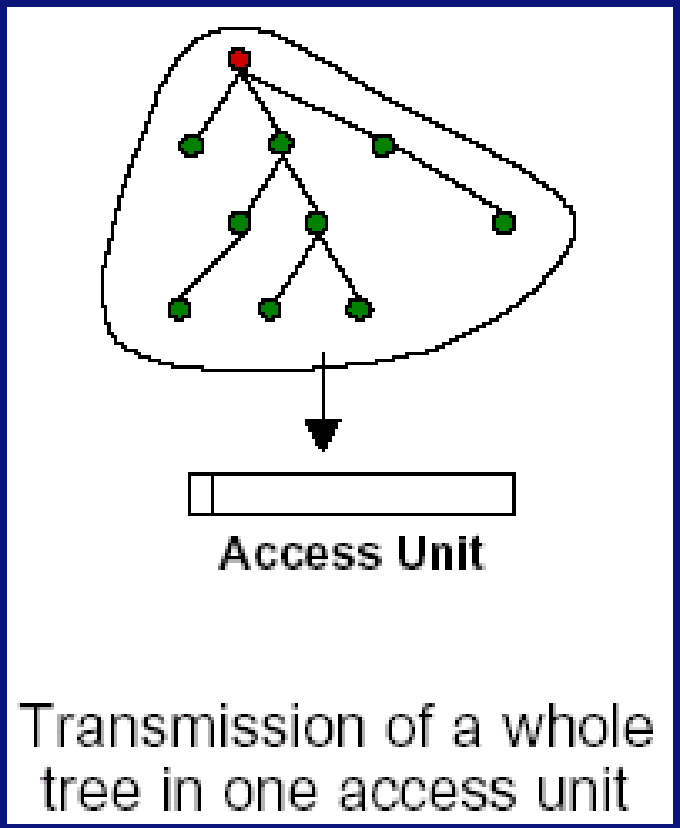
MPEG-7: part 1 Systems – flexible transmission

- **MPEG-7 Access Unit**
  - Contains a part of an MPEG-7 description
  - Smallest entity to which time stamps can be attached
  - Can be mapped into different transport layers
- **Finer granularity**
  - More than one sub-tree may need the same time
  - Access Units may contain more than one sub-tree

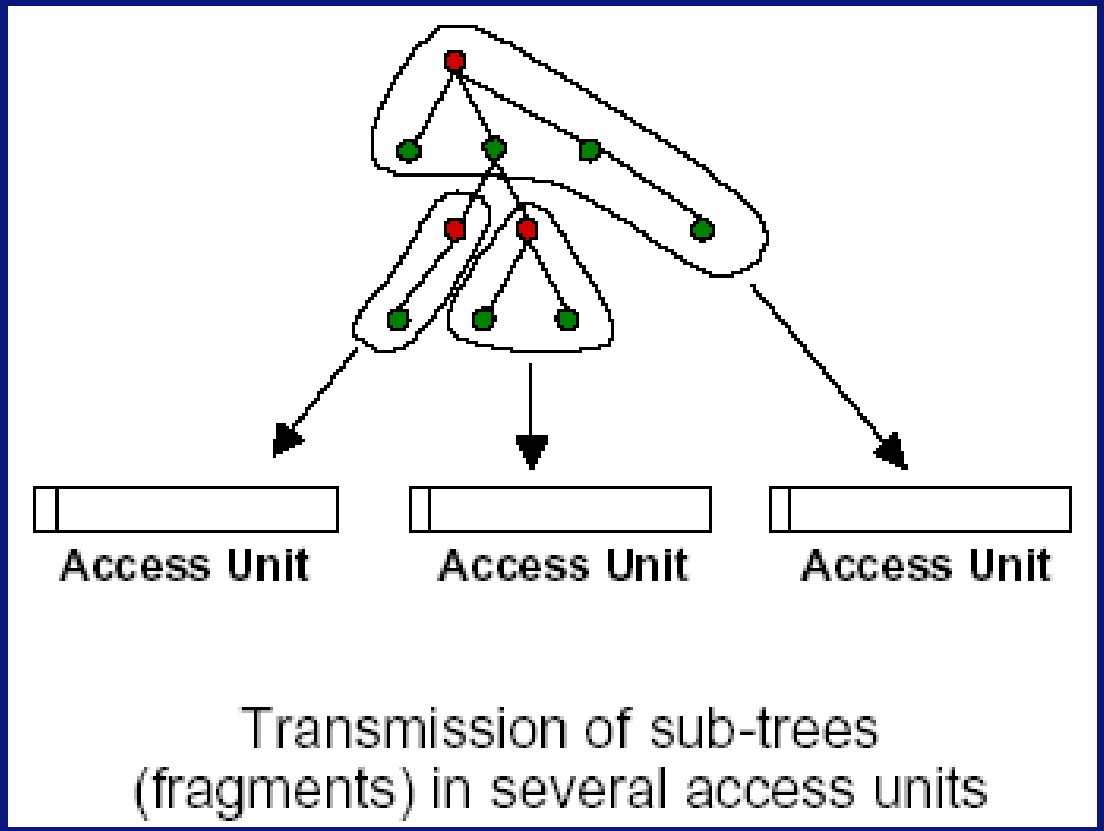
→ Sub-division of an Access Unit into several Fragment Update Units

===!"§==Systems=

MPEG-7: part 1 Systems – flexible transmission

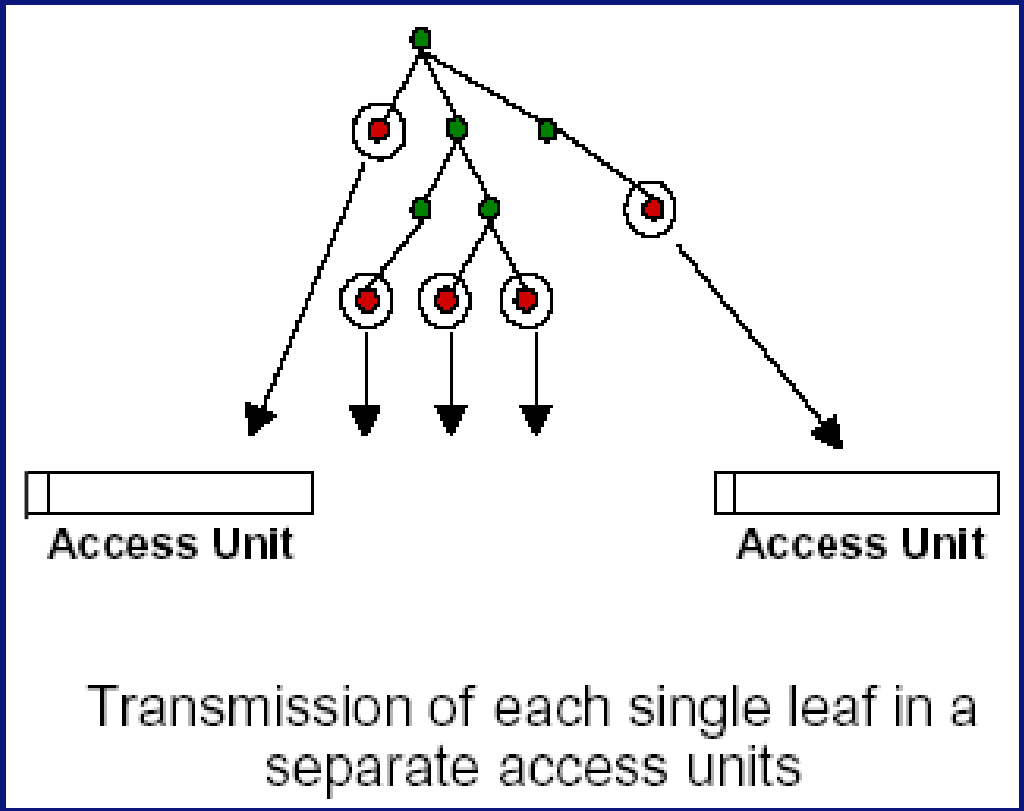


Transmission of a whole tree in one access unit



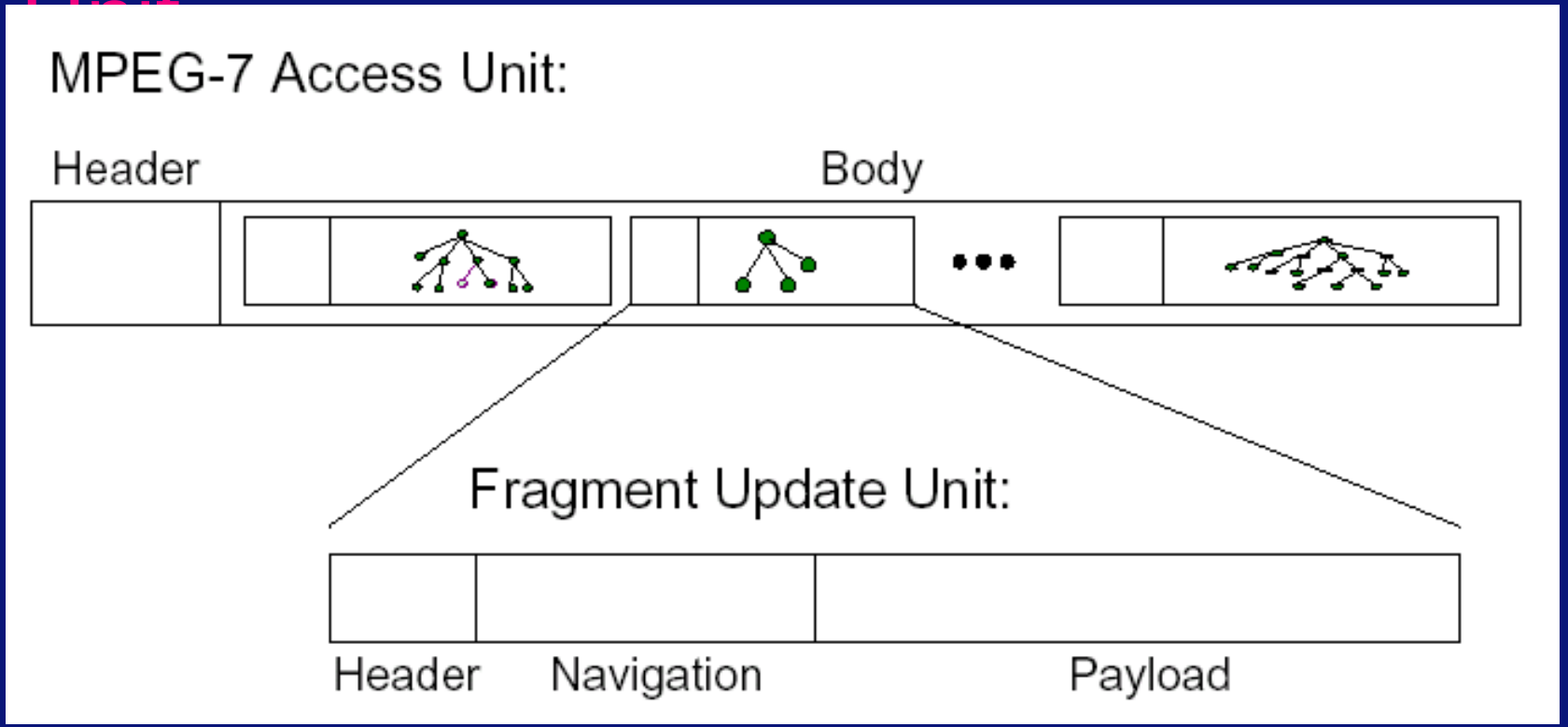
Transmission of sub-trees (fragments) in several access units

MPEG-7: part 1 Systems – flexible transmission



Transmission of each single leaf in a separate access units

MPEG-7: part 1 Systems – Access Unit

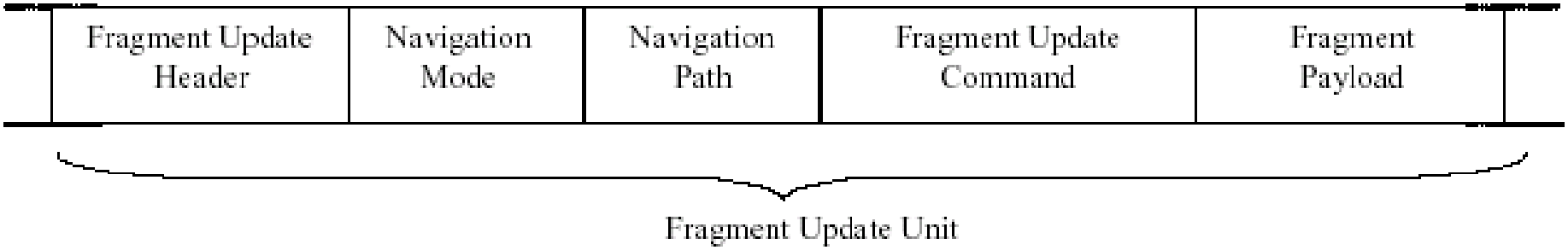


- Contains a part of an MPEG-7 description
- Smallest entity to which time stamps can be attached
- Can be mapped into different transport layers

===!"§==Systems=

MPEG-7: part 1 Systems – Fragment Update Unit

Fragment Update Unit - some more insight

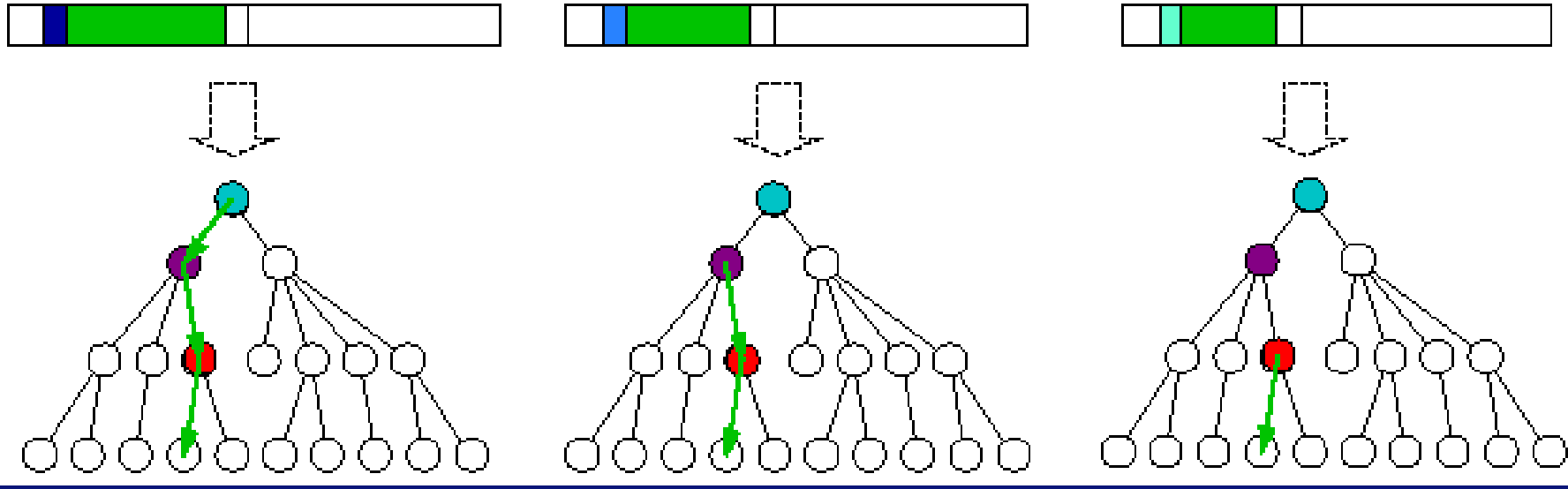


Fragment Update header:	parameters to decode fragment unit
Navigation mode:	specifies kind of addressing
Navigation path:	specifies path to the destination node
Fragment update command:	specifies command: add, change, delete
Fragment payload:	sub-tree (fragment) data

MPEG-7: part 1 Systems – navigation in description trees

Navigation in description trees

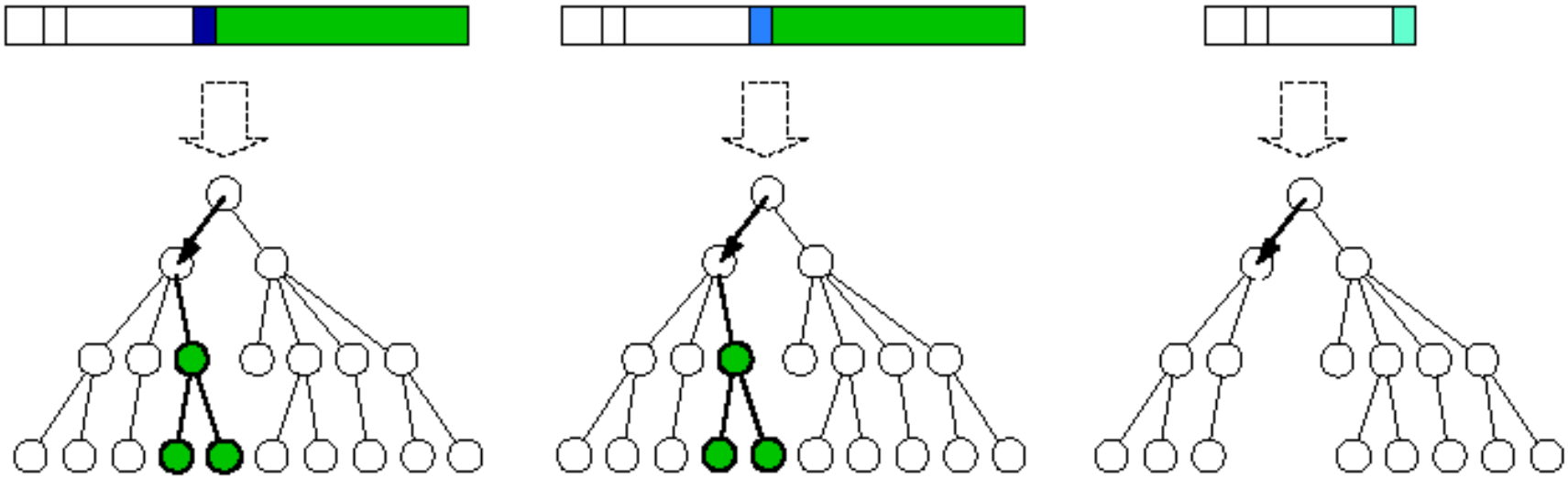
- Absolute wrt. the root node ( )
- Absolute wrt. the top-level node ( ) of the current DS
- Relative wrt. the current node ( )



MPEG-7: part 1 Systems – update commands

Dynamic update by fragment update commands

- **Add**: add a sub-tree at the current node
- **Change**: change a sub-tree at the current node
- **Delete**: delete sub-tree at the current node



MPEG-7: part 1 Systems – BiM

The Binary Format for MPEG-7 Descriptions allows...

- the bandwidth efficient binary transmission or storage,
- the dynamic update and the flexible streaming and
- searching or filtering on bitstream level

...of MPEG-7 DS instances

BiM basic features

- High compression ratio for structure up to 98%
- Specific datatype encoding leads to general compression of about 80%
- Generic XML encoding tool allowing extension of MPEG-7 DS‘s

=====!"§==Systems=

MPEG-7: part 1 Systems – BiM <-> XML

Relation to MPEG-7 textual format (XML)

· A bi-directional mapping between the BiM and the textual format is provided

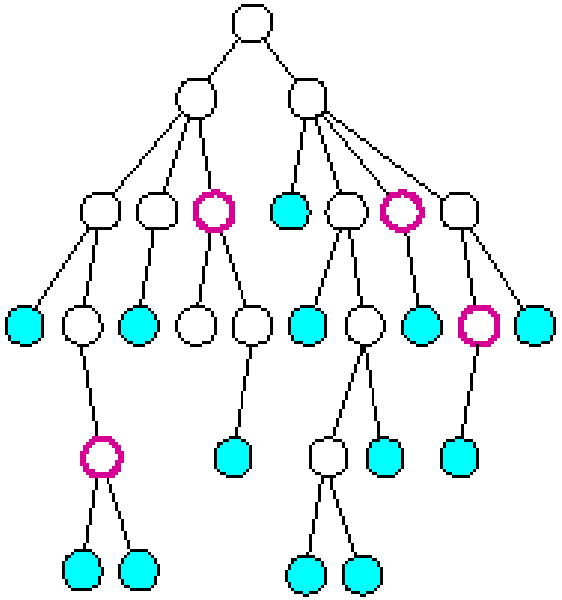
· When encode an MPEG-7 XML-file into BiM and then back to
MPEG-7 XML, both XML files are „canonically“ equivalent
– XML canonical format is specified by W3C
– Order of elements preserved
– Order of attributes ignored
– White spaces and XML comments are not preserved

→ The transmission of an MPEG-7 XML file and of the corresponding BiM stream result in identical

MPEG-7: part 1 Systems – BiM compression

How does it work?

- BiM is „schema aware“, i.e. the knowledge of the schema is exploited for efficient encoding



MPEG-7 DS

- Path through binary tree is encoded
- Assign codes to optional elements (○)
- Mandatory elements need no codeword

- Only leaf nodes (●) contain data
- Specific datatype encoding is applied

MPEG-7: part 1 Systems – conclusion

The MPEG-7 Systems tools allow

· flexible transmission and dynamic update of MPEG-7 Descriptions in

  arbitrary order

· choice between textual and binary format

· bandwidth efficient transport and storage of MPEG-7 Descriptions

  using MPEG-7 BiM

· searching and filtering of MPEG-7 Descriptions on textual and on binary level

===!"§==Systems=

MPEG-7: part 2 DDL

• specifies the language for defining the standard set of description tools (DSs, Ds, and datatypes) and for defining new description tools

• defines Description Schemes, Descriptors, and Datatypes by specifying their syntax, and allows their extension

- Descriptors (D): representations of Features, that define the syntax and the semantics of each feature representation

- Description Schemes (DS), that specify the structure and semantics of the relationships between their components, which may be both Ds and DSs

- A Description Definition Language (DDL), to allow the creation of new DSs and, possibly, Ds and to allow the extension and modification of existing DSs

MPEG-7: part 2 DDL

XML Schema:

- Structure
- Datatypes: Simple and Complex types
- Elements
- Inheritance
- Abstract types

MPEG-7 extensions:

- Array and Matrix datatype
- Enumerated datatypes for MimeType, CountryCode, RegionCode, CurrencyCode and CharacterSetCode
- Typed references

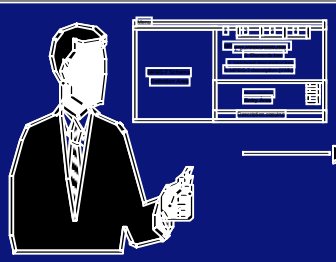
MPEG-7 part 2 DDL - example

MPEG-7 Schema Definition (DDL)

```
<complexType name="StateTransitionModelType"
   base="mpeg7:FiniteStateModelType" derivedBy="extension">
  <element name="Initial"
     type="mpeg7:ProbabilityVectorType" minOccurs="0" />
<element name="Transitions" type="mpeg7:ProbabilityMatrixType"/>
   <sequence minOccurs="1" maxOccurs="unbounded">
     <element name="State" type="mpeg7:ModelStateType"/>
   </sequence>
</complexType>
```

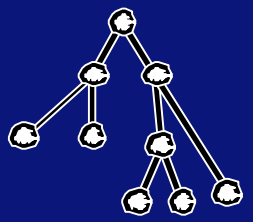
MPEG-7 Package

MPEG-7 Schema / Package Parser



User Input

MPEG-7 Visual Annotation Tool Interface



MPEG-7 Elements

MPEG-7 Description

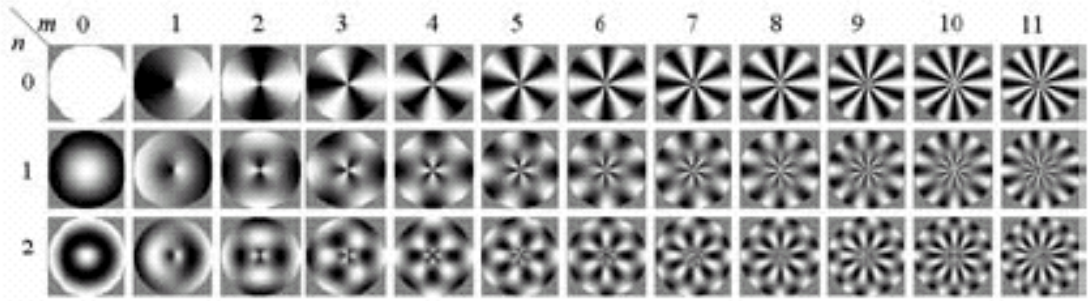
```
<StateTransitionModel numberStates="3">
  <Initial Size="3"> 0.25 0.5 0.25 </Initial>
  <Transitions Size1="3" Size2="3">
     0.2 0.2 0.6 0.1 0.8 0.1 0.3 0.3 0.4
   </Transitions>
  <State><Label> Pass </Label></State>
  <State><Label> Shot on goal </Label></State>
  <State><Label> Goal score </Label></State>
</StateTransitionModel>
```

MPEG-7 part 2 DDL example: MPEG-7 Visual Descriptor Definition (Region Shape)

- MPEG-7 Region Shape (DDL):

```
<complexType name="RegionShapeType" final="#all">
  <complexContent>
   <extension base="mpeg7:VisualDType">
     <element name="ArtDE">
         <simpleType>
           <restriction base="mpeg7:listOfUnsigned4Type>
             <length value="35"/>
           </restriction>
         </simpleType>
     </element>
   </extension>
  </complexContent>
</complexType>
```

Angular Radial Transform basis functions (**35** dimensions)



...short break...

Thank you for your attention so far !!!

<u>After the break Session 2 will continue with:</u>

- structure of the standard –parts 3-8 of MPEG-7
- links and hints on MPEG-7 activity

===!"§==Systems=

Structure of MPEG-7: part 3 Visual

Visual Features

Color

ColorFeatureDescriptors

DominantColor
ScalableColor
ColorLayout
ColorStructure
GofGopColor

ColorSupporting Tools

ColorSpace
ColorQuantization

Basic Structures

DescriptorContainers
GridLayout
TimeSeries
MultipleView

Basic Supporting Tools

TemporalInterpolation
Spatial2DcoordinateSystem

Texture

HomogeneousTexture
TextureBrowsing
EdgeHistogram

Shape

RegionShape
ContourShape
Shape3D

Motion

CameraMotion
MotionTrajectory
ParametricMotion
MotionActivity

Localization

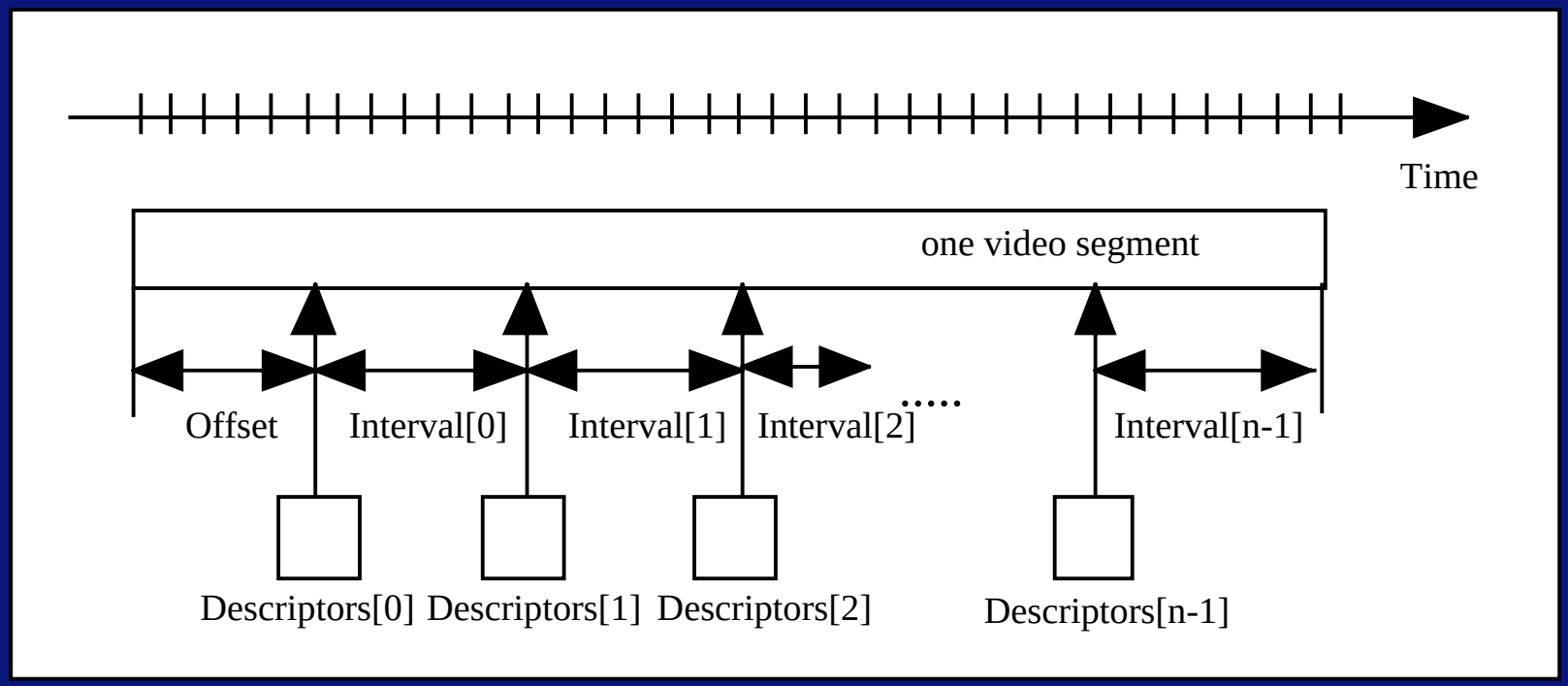
RegionLocator
SpatioTemporalLocator

Other

FaceRecognition

===!"§==Systems=

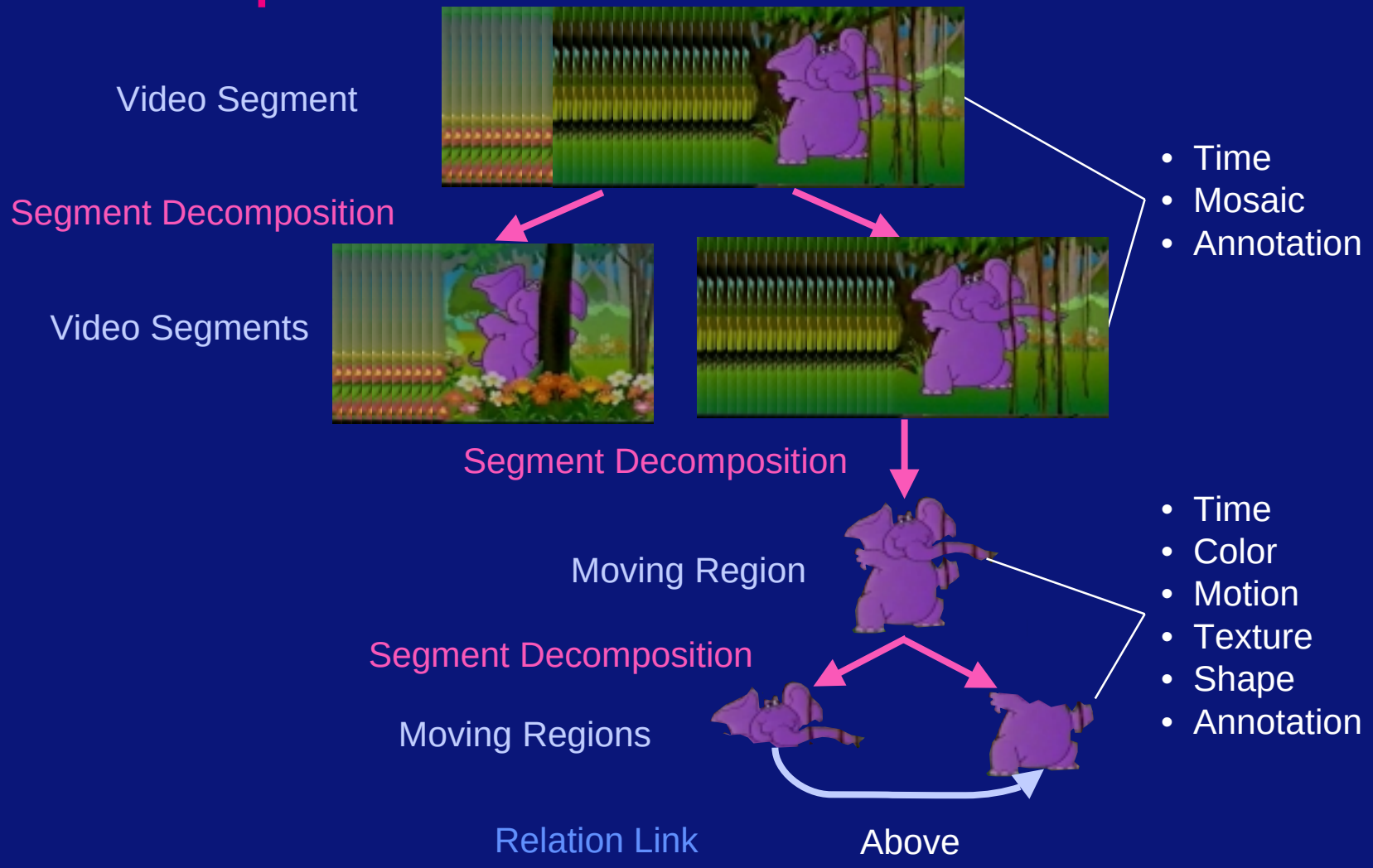
MPEG-7: part 3 Visual



- segmentation is one of the basic concepts
- irregular (nonlinear) time series consisting of segments and their feature description
- a bunch of models is defined for transformations of coordinate systems, color spaces, texture etc.

===!"§==Systems=

MPEG-7: part 3 Visual - structure description



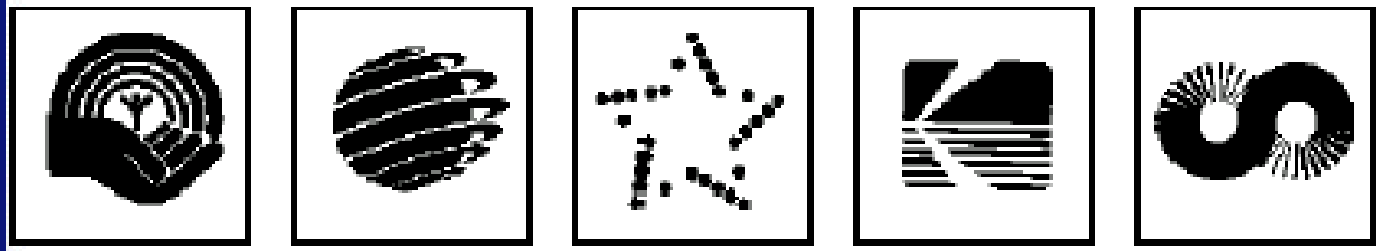
Ana Belén Benítez, Columbia University

MPEG-7 part 3 Visual - example (structure desc.)

```
<VideoSegment id = “RootVS">
        <SegmentDecomposition gap = ”false" overlap = ”false” decompositionType = "temporal">
        <VideoSegment id = “PinkyFlowersVS"> … </VideoSegment>
        <VideoSegment id = “PinkyVS">
            <MediaTime> <MediaTimePoint> T0:1:30 </MediaTimePoint>
            <MediaDuration> PT6M17S </MediaDuration> </MediaTime>
            <SegmentDecomposition gap = ”true" overlap = ”false”  decompositionType = ”spatial">
                <MovingRegion id=“PinkyMR”> ... </MovingRegion>
                        <TextAnnotation> <FreeText> Elephant Pinky </FreeText> </TextAnnotation>
                    <SegmentDecomposition gap = ”true" overlap = ”false” decompositionType = ”spatial">
                        <MovingRegion id=“HeadMR”> …
                            <RelationLink name=“aboveOf” target=“BodyMR”/>
                        </MovingRegion>
                        <MovingRegion id=“BodyMR”> … </MovingRegion>
                    </SegmentDecomposition>
                </MovingRegion>
            </SegmentDecomposition>
        </VideoSegment>
    </SegmentDecomposition>
```

MPEG-7 part 3 Visual: Shape Descriptors

- 2D contour based descriptors:
  - Rotation and scale invariance and invariance to small non-rigid deformations
- 2D area based descriptors:
  - Complex shapes such as trademarks

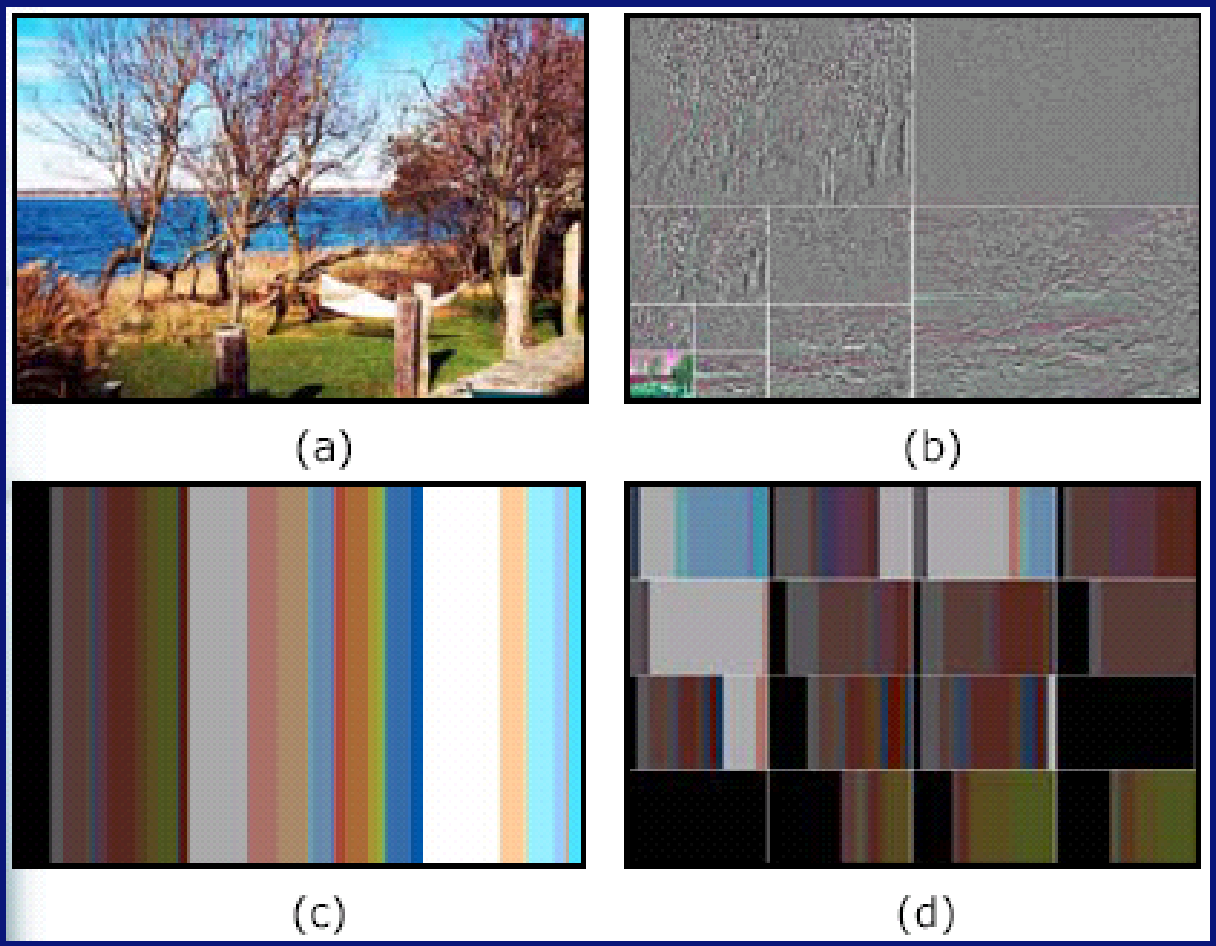


- 3D shape:



MPEG-7 part 3 Visual - example (Global Image Features)

Example: Global Image Features



a. Source Image

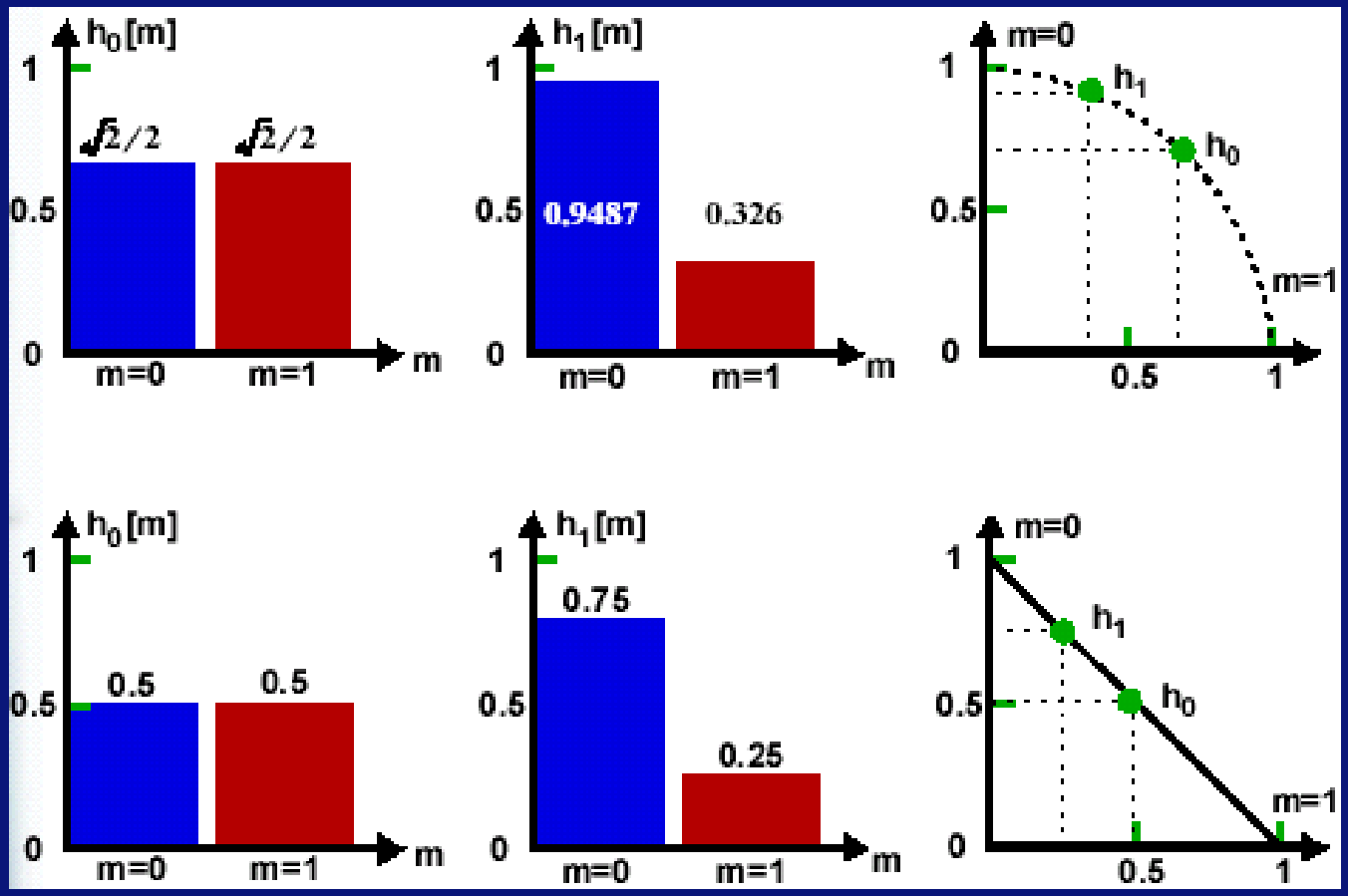
b. Texture:
Spatialfrequency energy
(ca.10 dimensions)

c. Color: color histogran
(ca. 100 dimensions)

d. Color composition:
grid color histogram
(ca. 100 dimensions)

- **Non-normative**:
  – Histogram normalizati
  – Distance metrics

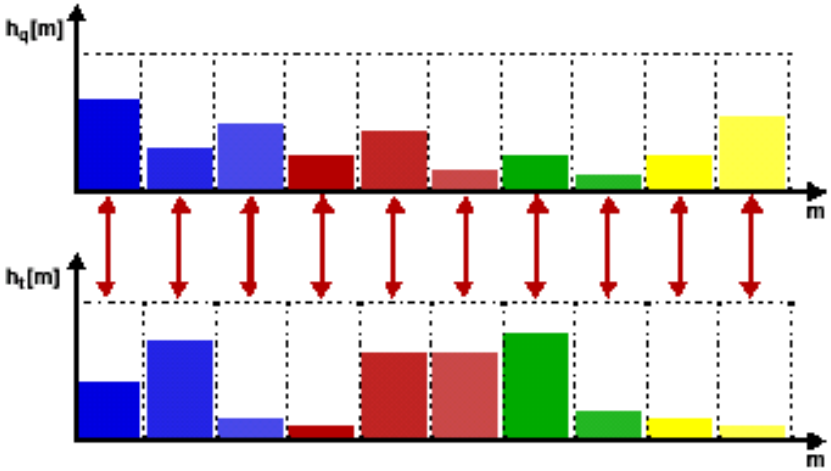
MPEG-7 part 3 Visual – example (Histogram Normalization)

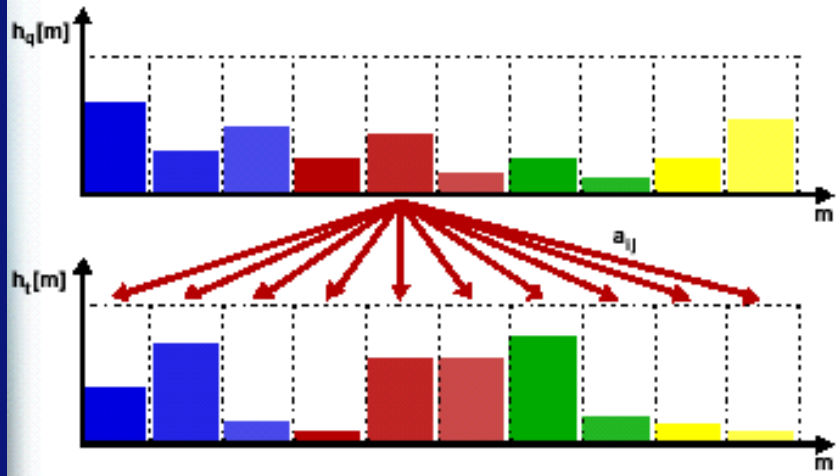


- Removes dependency on image size (#samples)
- L2 norm (Euclidean): vectors on surface of Hypersphere
- L1 norm: vectors on surface of simplex **0**

MPEG-7 part 3 Visual – example (Color Histogram Matching)

- Minkowski Metrics (i.e., L1, L2)
- Direct comparison of histogram bins (like colors only)





- Quadratic Metrics
- Cross comparison of colors (color similarity or correlation)

Structure of MPEG-7: part 4 Audio

1. Low Level Description Tools

   - AudioFramework

2. High Level Description Tools

   - Audio Signature Description Scheme

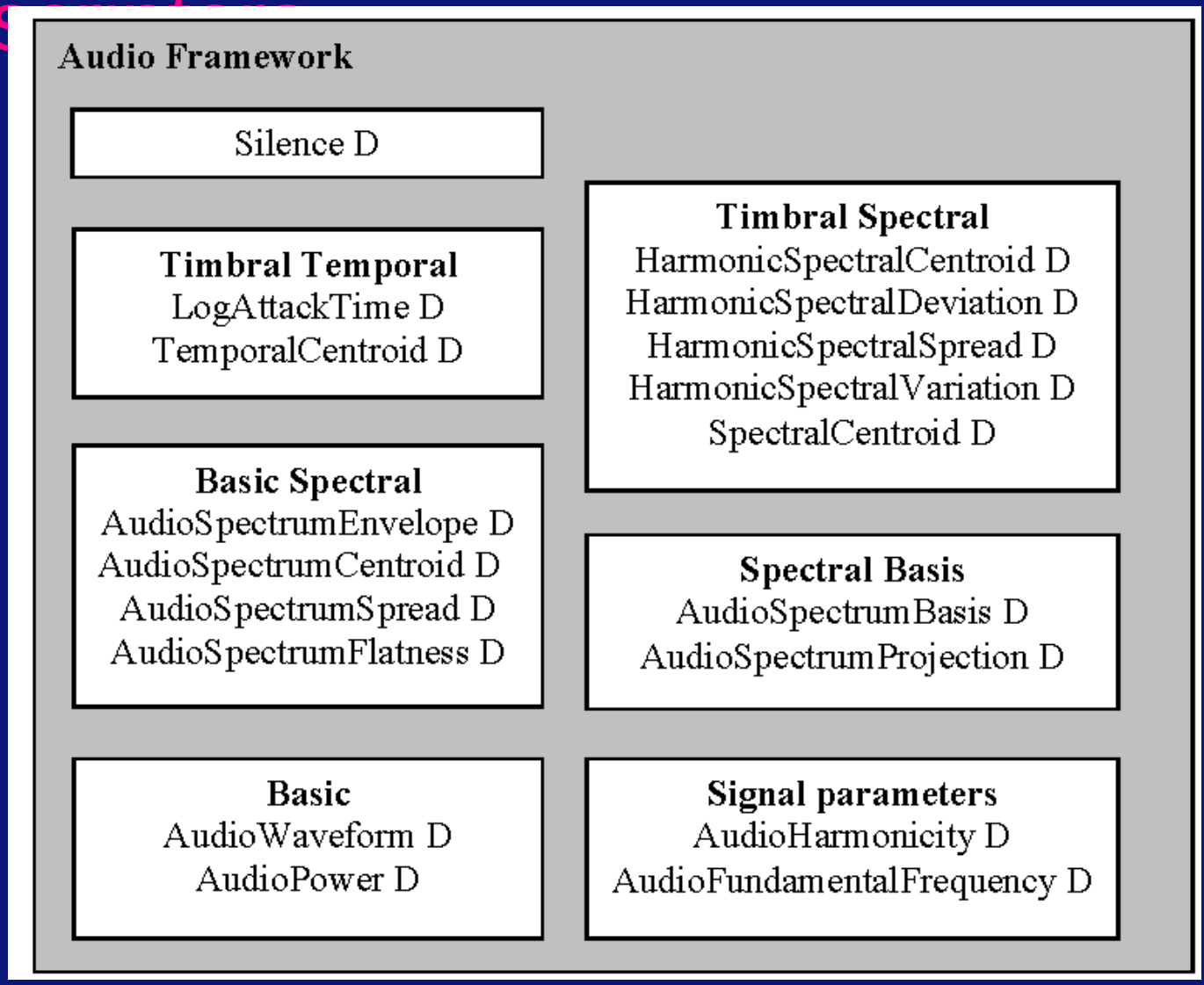
   - Musical Instrument Timbre Description Tools

   - Melody Description Tools

   - General Sound Recognition and Indexing Description Tools

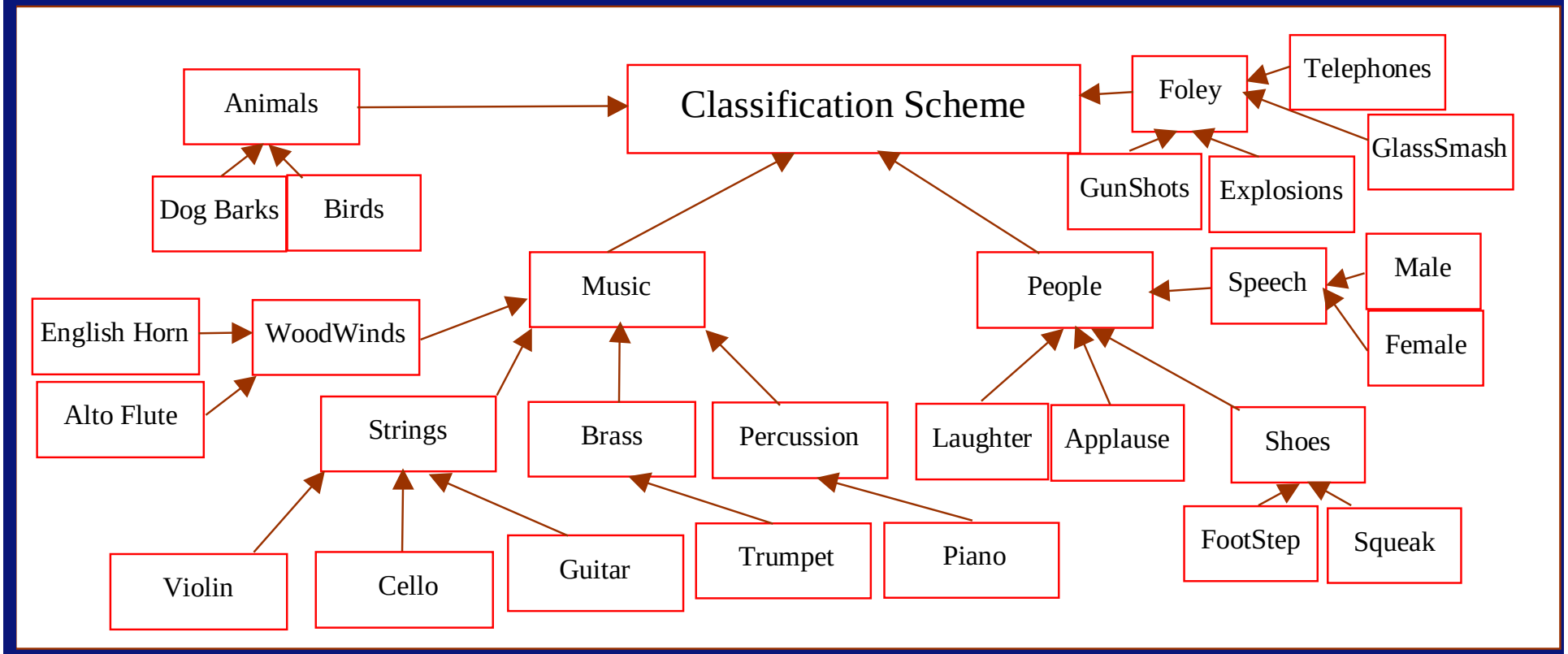
   - Spoken Content Description Tool

MPEG-7: part 4 Audio - low level descriptors



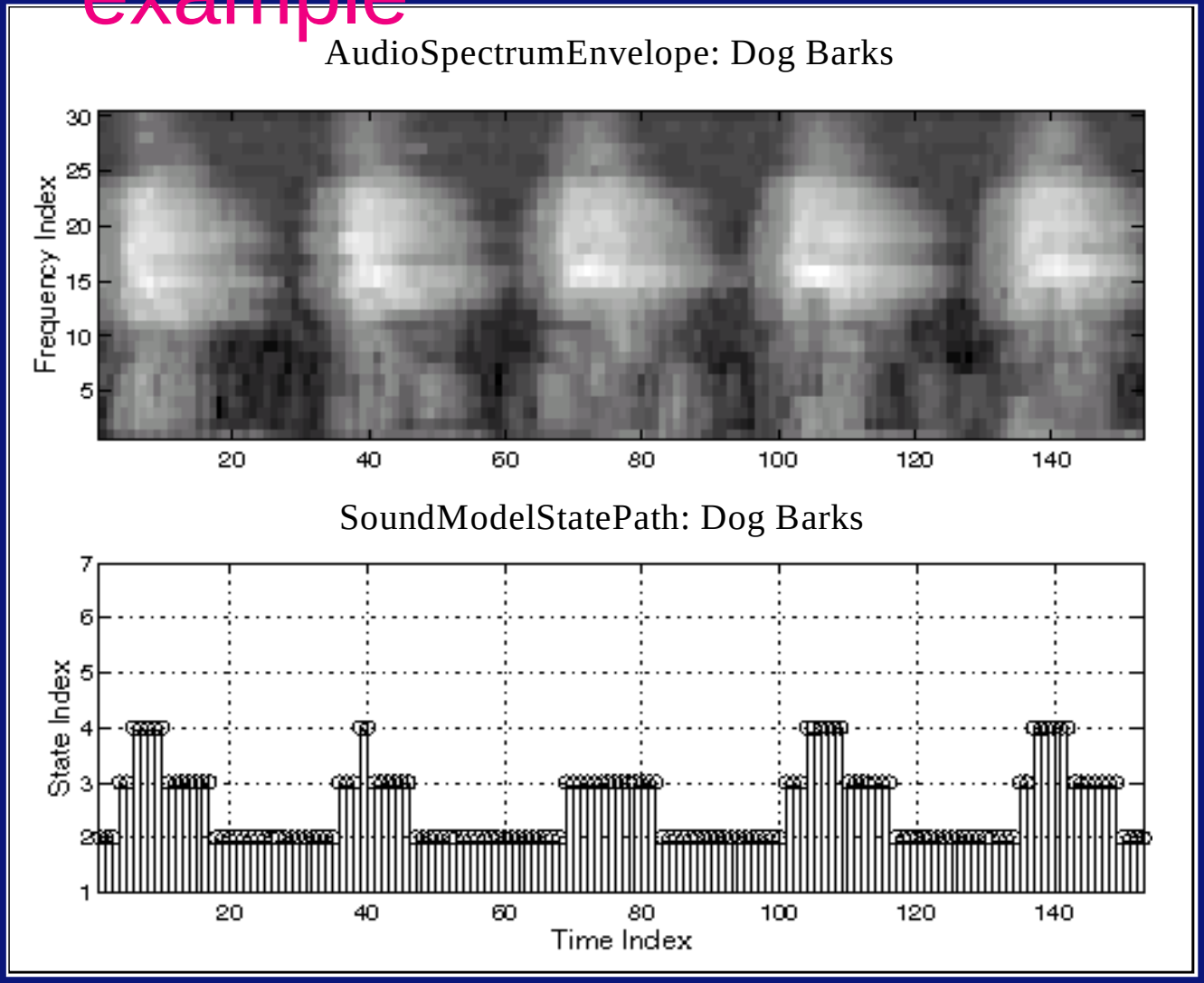
MPEG-7: part 4 Audio - high level example

General Sound Recognition and Indexing



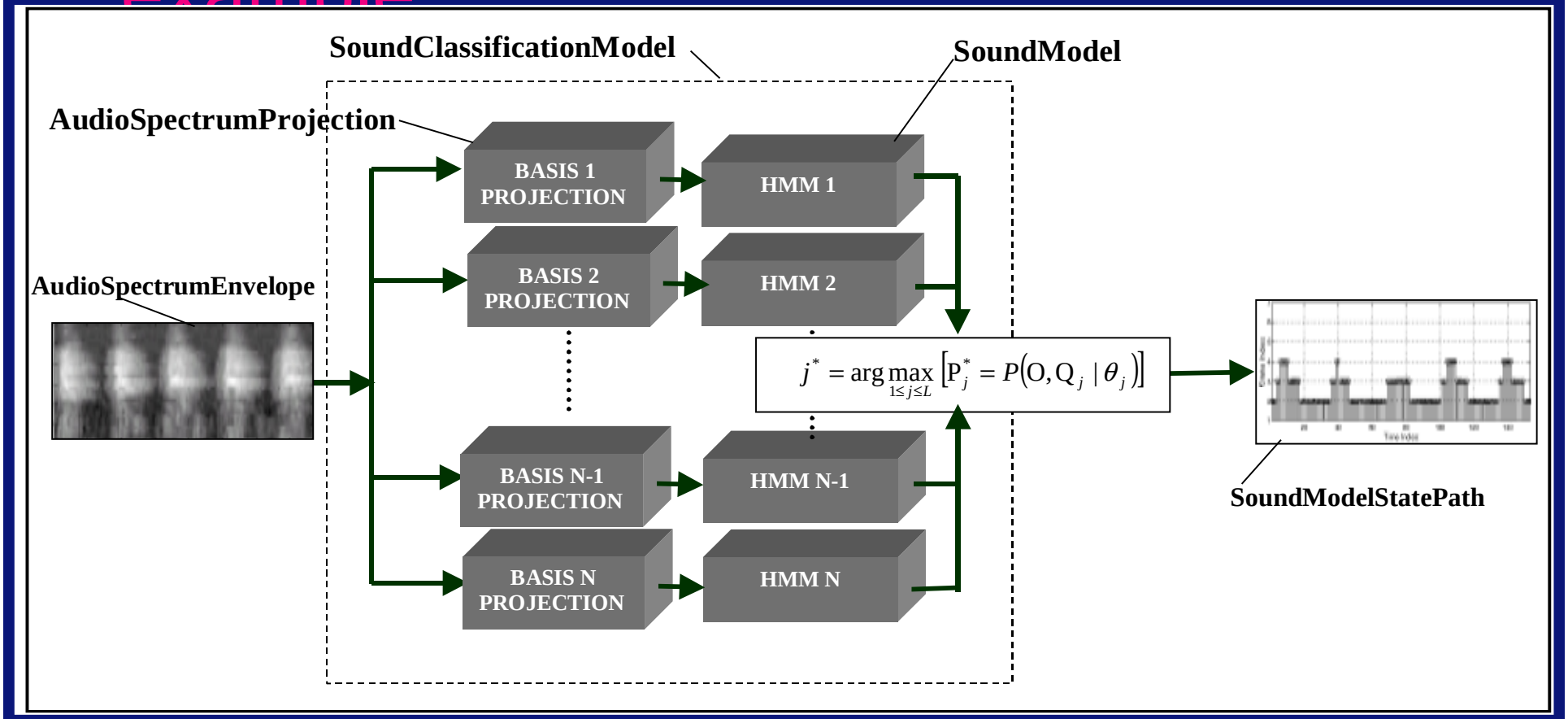
Example Classification Scheme for General Audio Content

MPEG-7: part 4 Audio - high level example



Spectrum of a dog barking and and the state-path generated by a SoundModel

MPEG-7: part 4 Audio - high level example



Use of multiple HMM models for automatic classification and indexing of audio content

MPEG-7: part 4 Audio - high level example

spoken content Description Tools:

Recall of audio/video data by memorable spoken events. Given a film or video recording where a character or person spoke a particular word or sequence of words, source media would be known, and the query would return a position in the media.

Spoken Document Retrieval. Given a database consisting of separate spoken documents - the result of the query is the relevant documents, optionally the position in those documents of the matched speech

Annotated Media Retrieval. result of the query is the media which is annotated with speech, not the speech itself. An example is a photograph retrieved using a spoken annotation.

MPEG-7: part 4 Audio - high level example

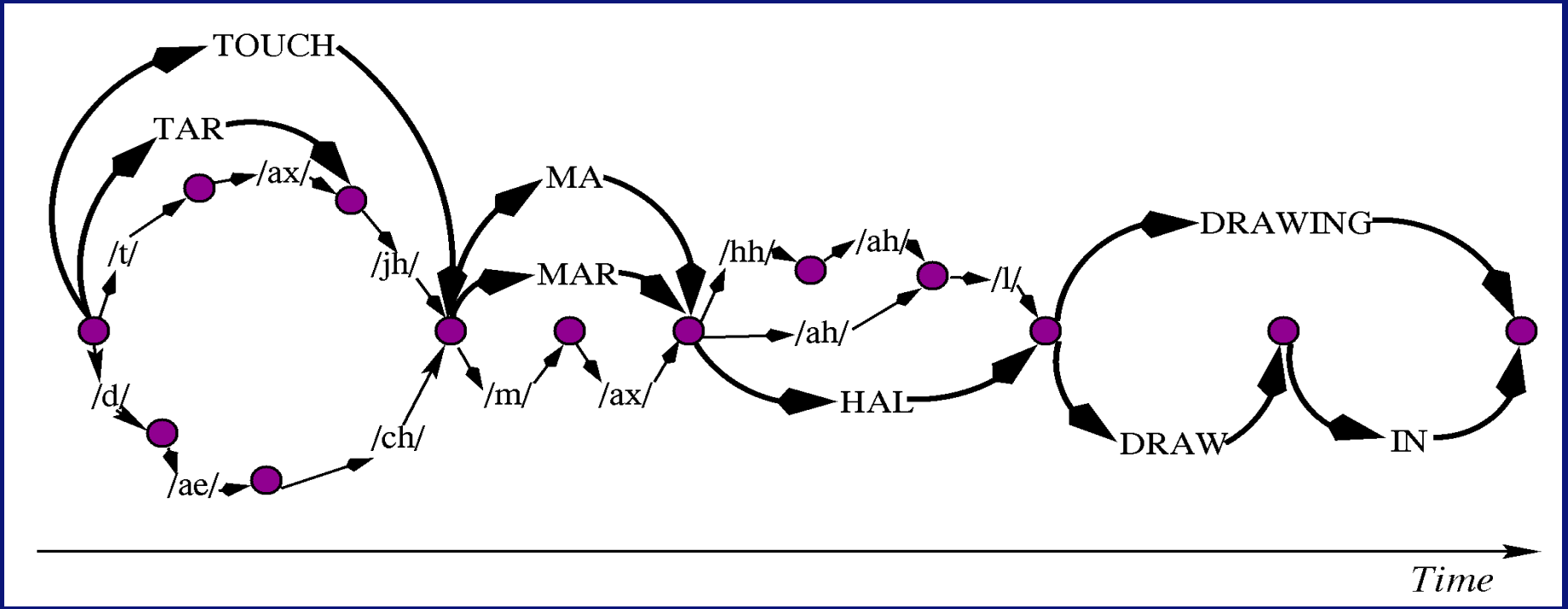
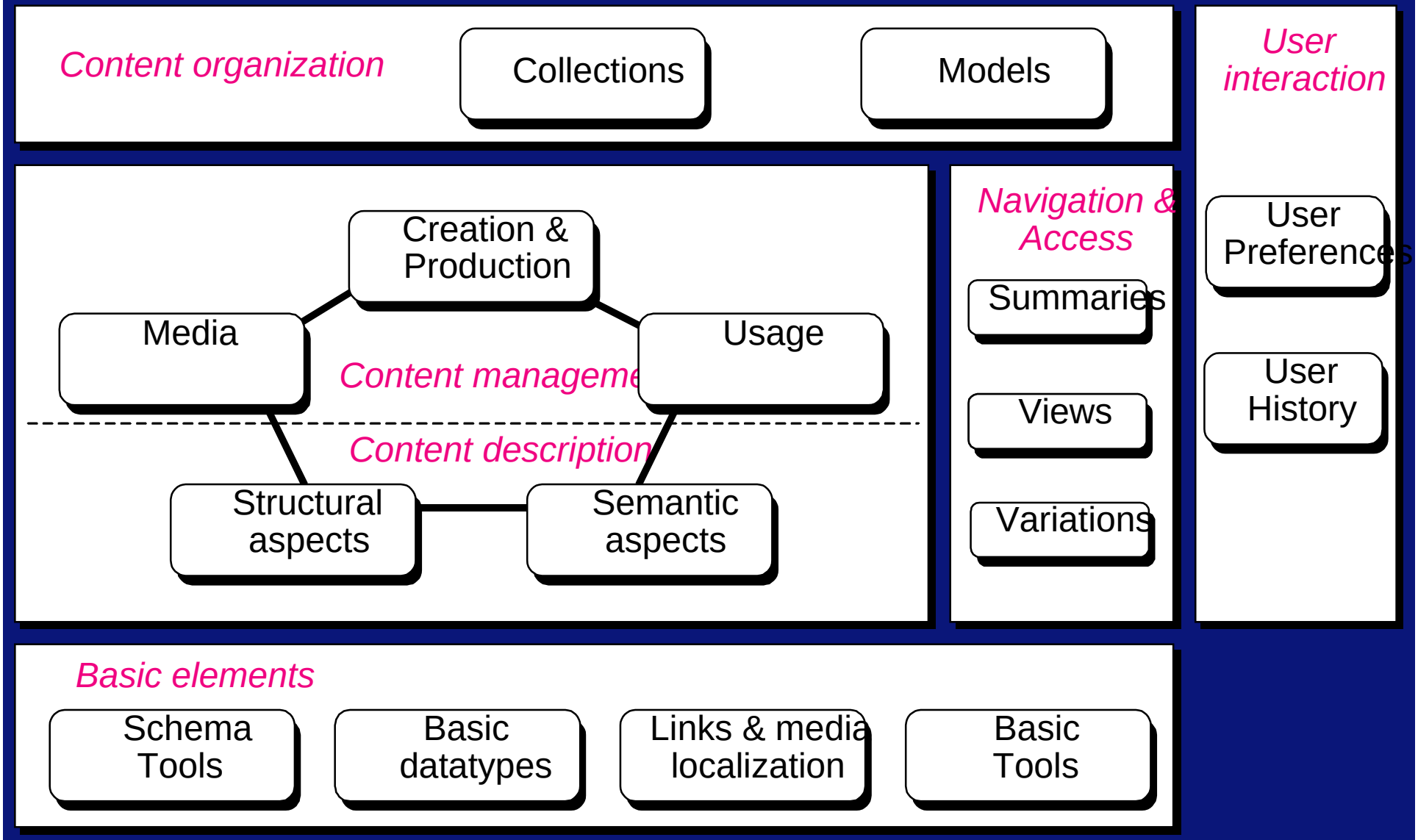


Figure : A lattice structure for an hypothetical (combined phone and word) decoding of the expression “Taj Mahal drawing …”. It is assumed that the name ‘Taj Mahal’ is out of the vocabulary of the ASR system

Structure of MPEG-7: part 5 MDS

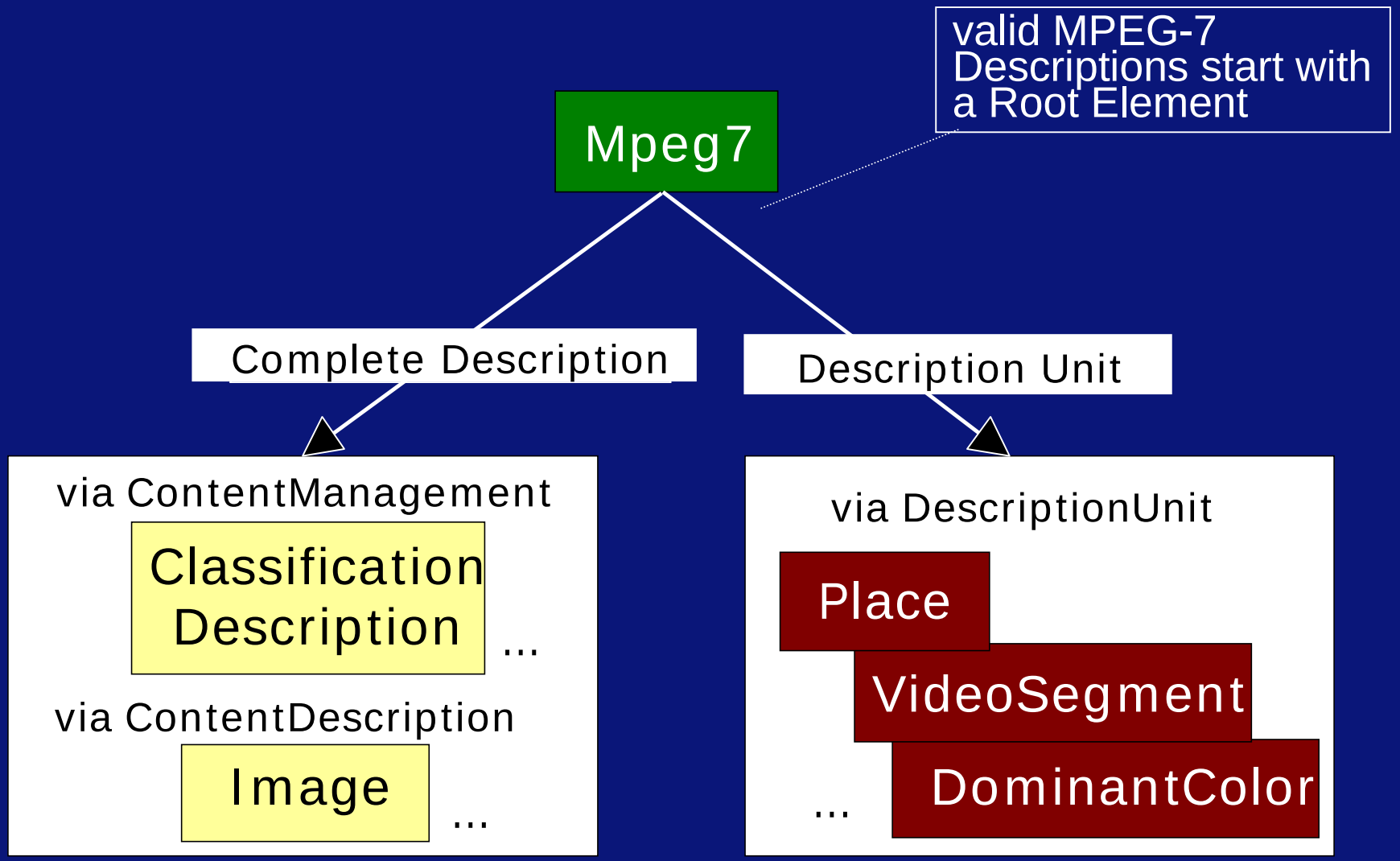


MPEG-7: part 5 MDS

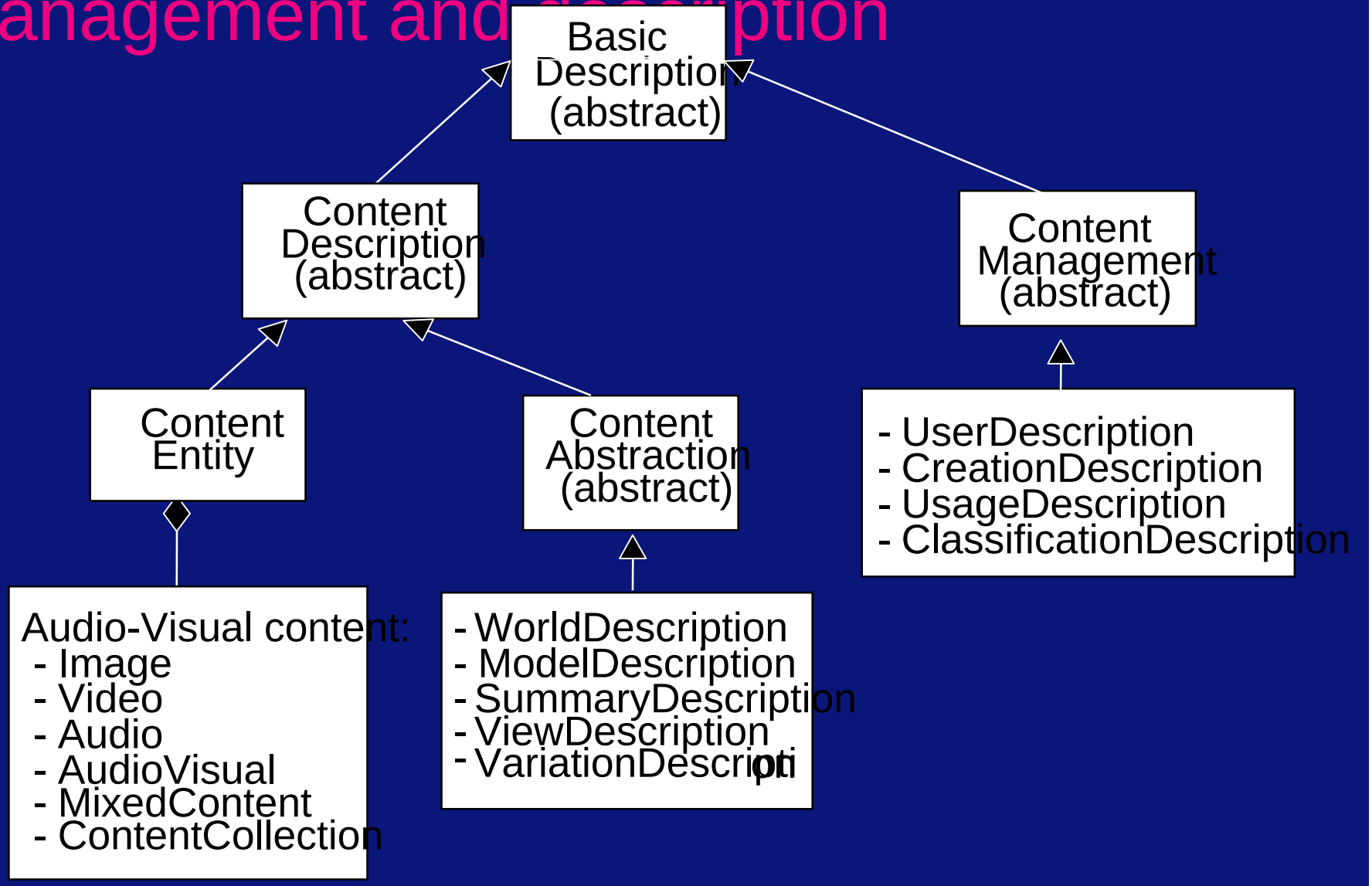
Complete Description: A complete description of multimedia content using the top-level elements (the Content Management and the Content Description). For example, the full description of an image is a complete description. The definition of "complete" will depend on the application using the description.

Description Unit: A description consisting of an arbitrary descriptor or description scheme instance. Unlike a complete description which will usually contain a "semantically-complete" description, a description unit can be used to send a partial description as required by an application – such as a description of a place, a shape and texture descriptor and so on. A description unit is enclosed by the top-level element called DescriptionUnit.

MPEG-7: part 5 MDS



MPEG-7: part 5 MDS content management and description



MPEG-7: part 5 MDS content management and description

The following example gives an MPEG-7description of the creation information for a sports video:

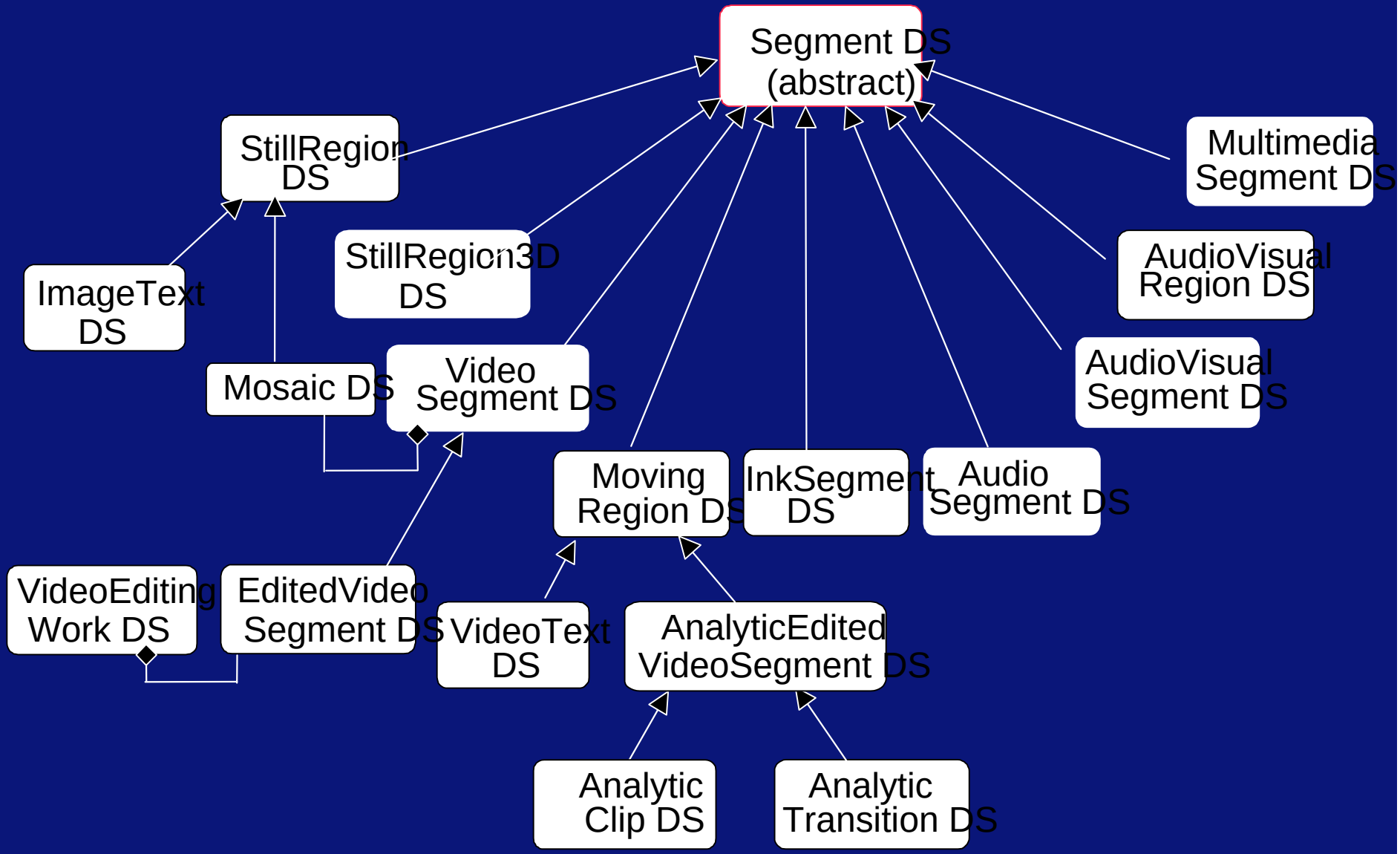
```
<Mpeg7>
  <Description xsi:type="CreationDescriptionType">
    <CreationInformation>
      <Creation>
        <Title type="popular">Subway series</Title>
        <Abstract>
          <FreeTextAnnotation> Game among city rivals
            </FreeTextAnnotation>
        </Abstract>
        <Creator>
          <Role href="urn:mpeg:mpeg7:cs:RoleCS:2001:PUBLISHER"/>
          <Agent xsi:type="OrganizationType">
            <Name>Sports Channel</Name>
          </Agent>
        </Creator>
      </Creation>
    </CreationInformation>
  </Description>
</Mpeg7>
```

MPEG-7 part5 MDS: Collection Model Example

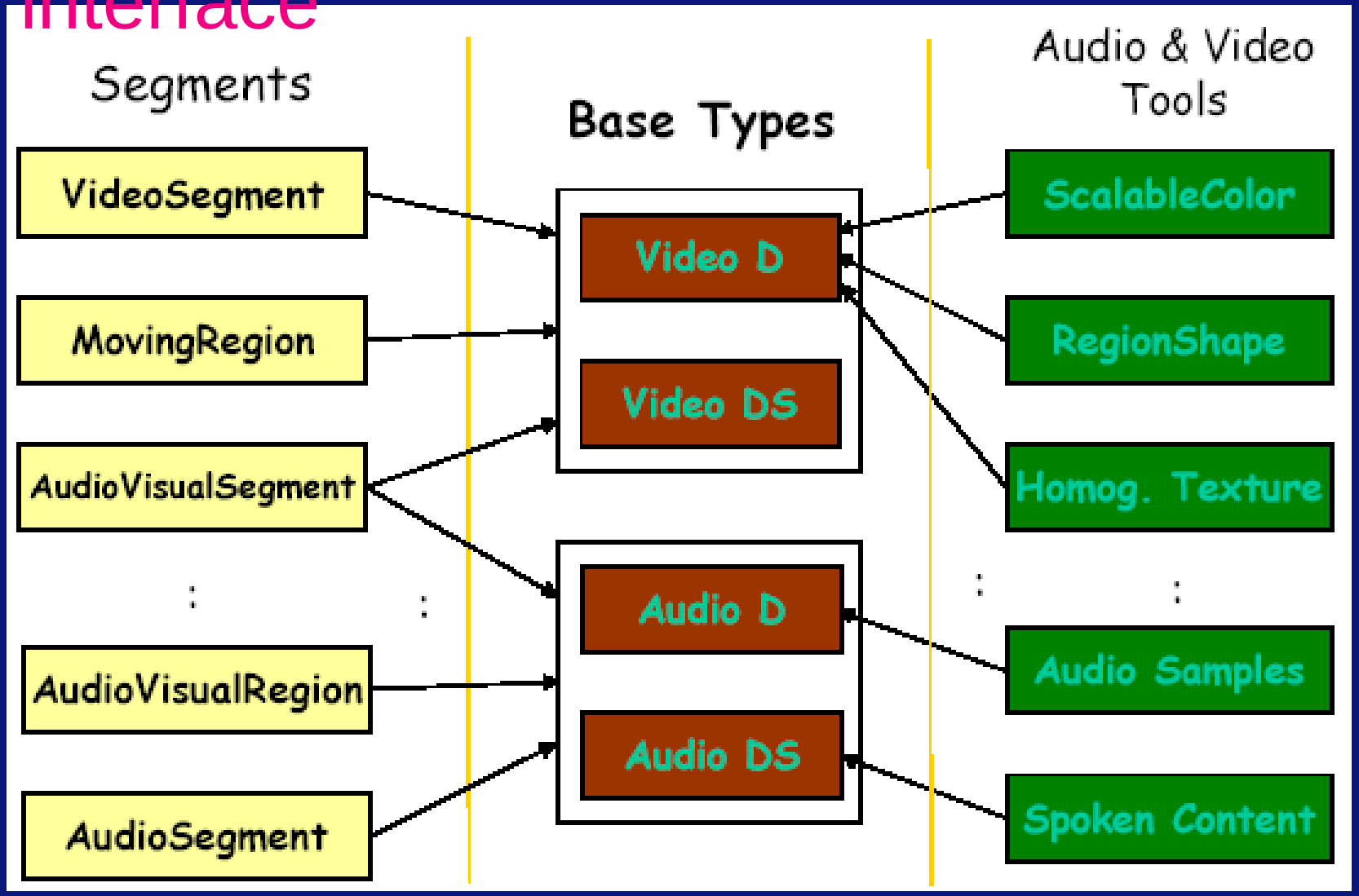
MPEG-7 description of a collection model of "sunsets" that contains two images depicting sunset scenes:

```
<Mpeg7>
 <Description xsi:type="ModelDescriptionType">
  <Model xsi:type="CollectionModelType" confidence="0.7
   reliability="0.5" function="described">
    <Label><Name> Sunsets </Name></Label>
     <Collection xsi:type="ContentCollectionType">
      <Content xsi:type="ImageType">
       <Image>
        <MediaLocator xsi:type="ImageLocator
         <MediaUri>sunset1.jpg</MediaUr
        </MediaLocator>
       </Image>
      </Content>
      <Content xsi:type="ImageType">
       <Image>
        <MediaLocator xsi:type="ImageLocator
         <MediaUri>sunset2.jpg</MediaUr
        </MediaLocator>
       </Image>
      </Content>
     </Collection>
   </Model>
 </Description>
</Mpeg7>
```

MPEG-7: part 5 MDS segmentation

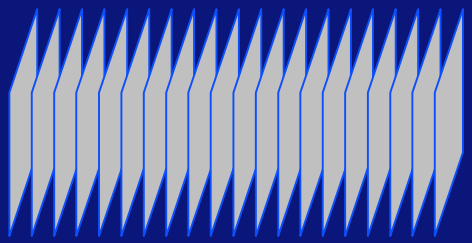


MPEG-7: part 5 MDS segment interface



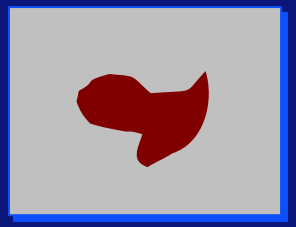
MPEG-7: part 5 MDS audiovisual segment descriptors

Video segments



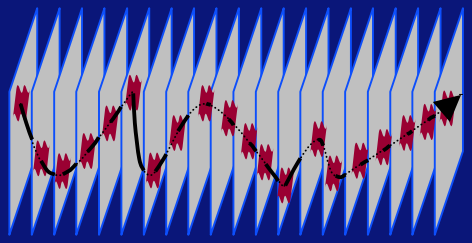
- Color
- Camera motion
- Motion activity
- Mosaic

Still regions



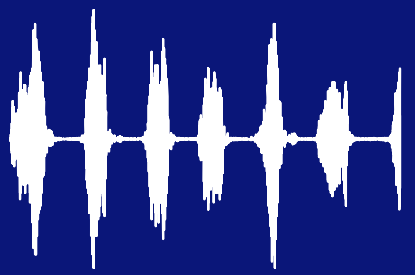
- Color
- Shape
- Position
- Texture

Moving regions



- Color
- Motion trajectory
- Parametric motion
- Spatio-temporal shape

Audio segments



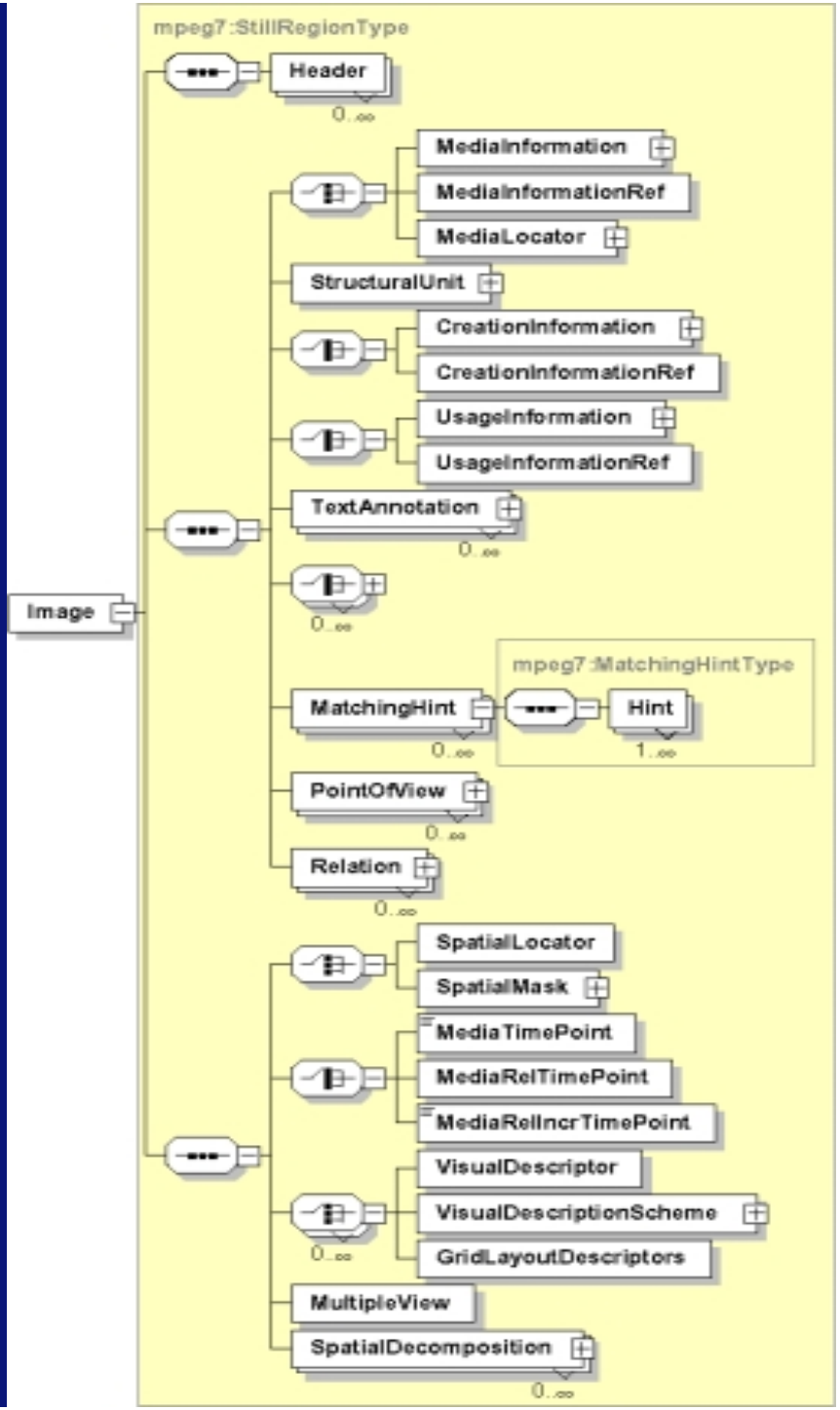
- Spoken content
- Spectral characteri-zation
- Music: timbre, melody

MPEG-7: part 5 MDS
e.g. Image Description



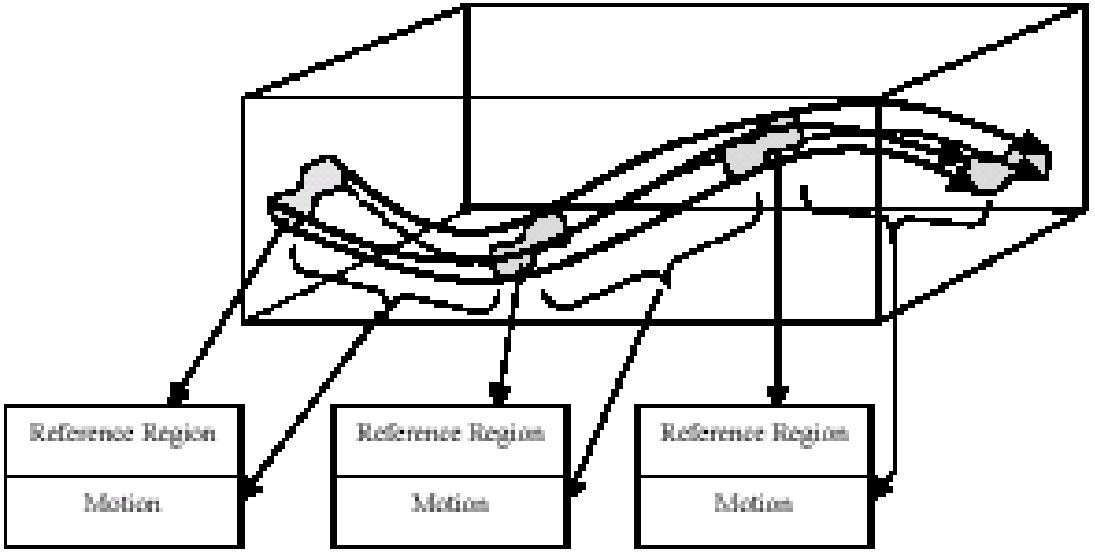
- Content management metadata:
  - Media information
  - Creation & production; usage
- Visual features:
  - Visual descriptors and DSs
- Structure:
  - Regions, spatial decomposition, spatio-temporal relationships

===!"§==Systems=



MPEG-7: part 5 MDS: Moving Regions



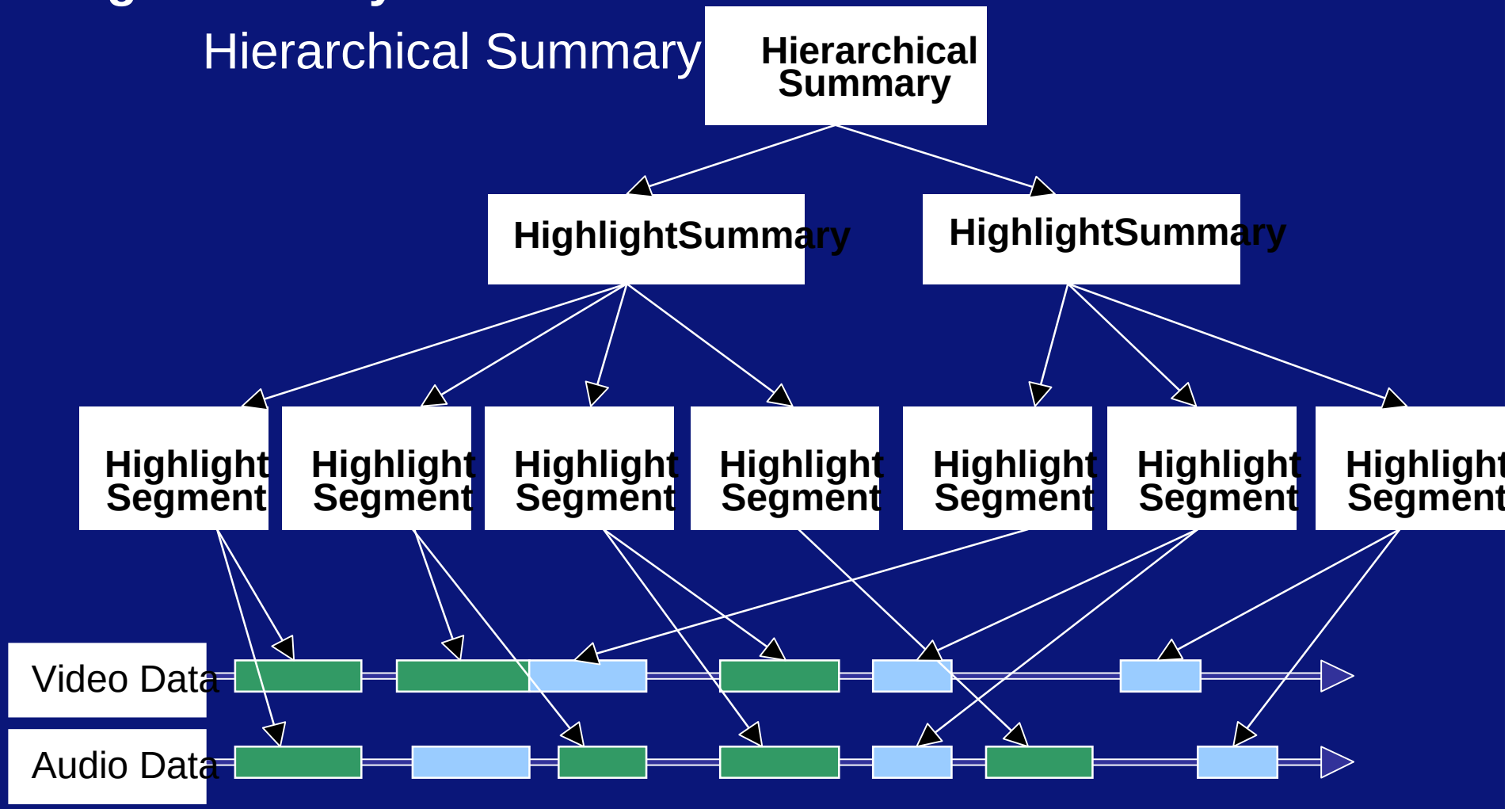


Moving Regions:
- Region Feature description (shape, color, texture)
- Spatio-Temporal Locator
- Regions Motion Trajectory and Tracking

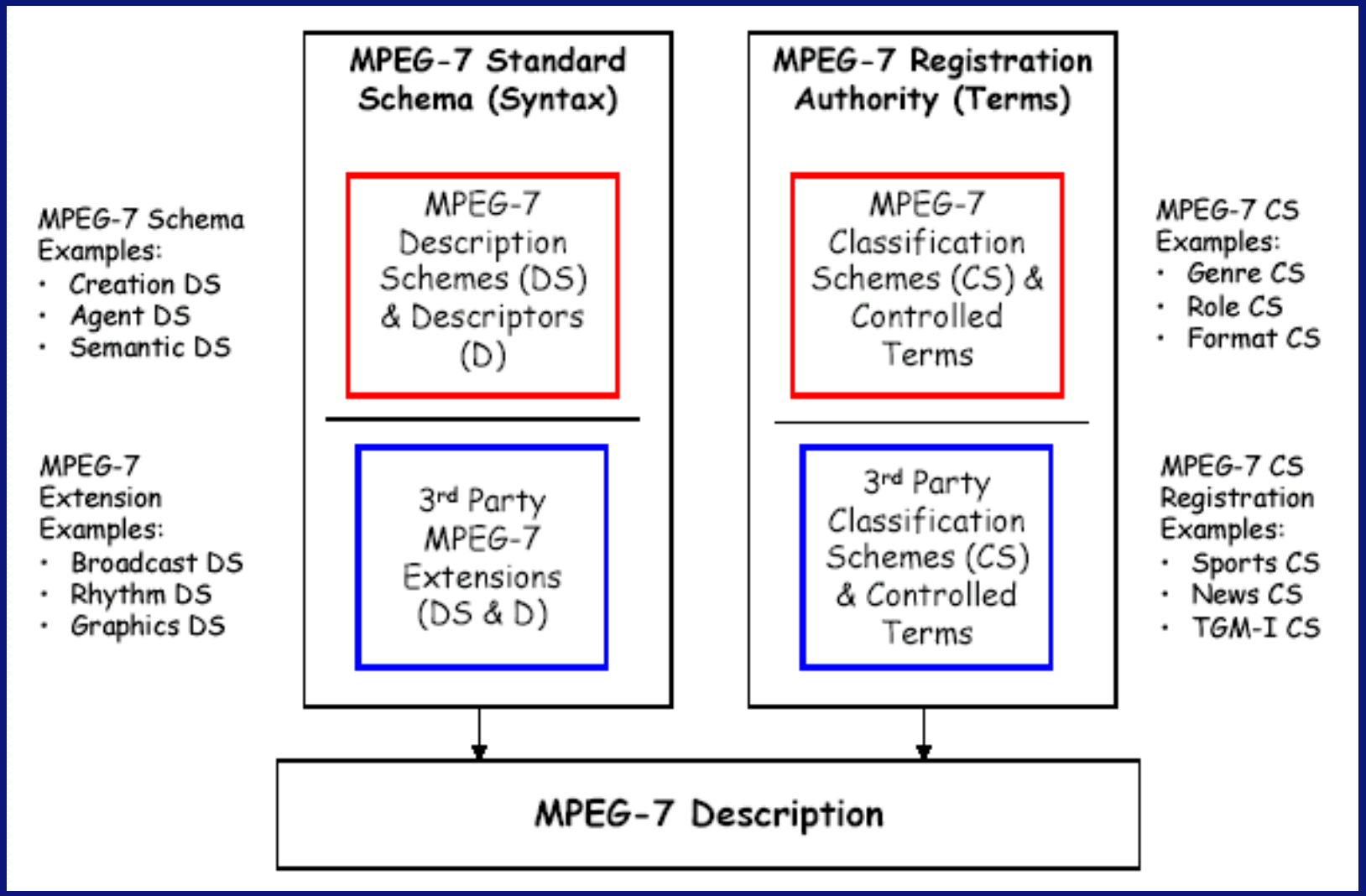
MPEG-7: part 5 MDS navigation and access

e.g. Summary DS:

Hierarchical Summary



MPEG-7: part 5 MDS Classification Schemes



MPEG-7: part 5 MDS Classification Schemes

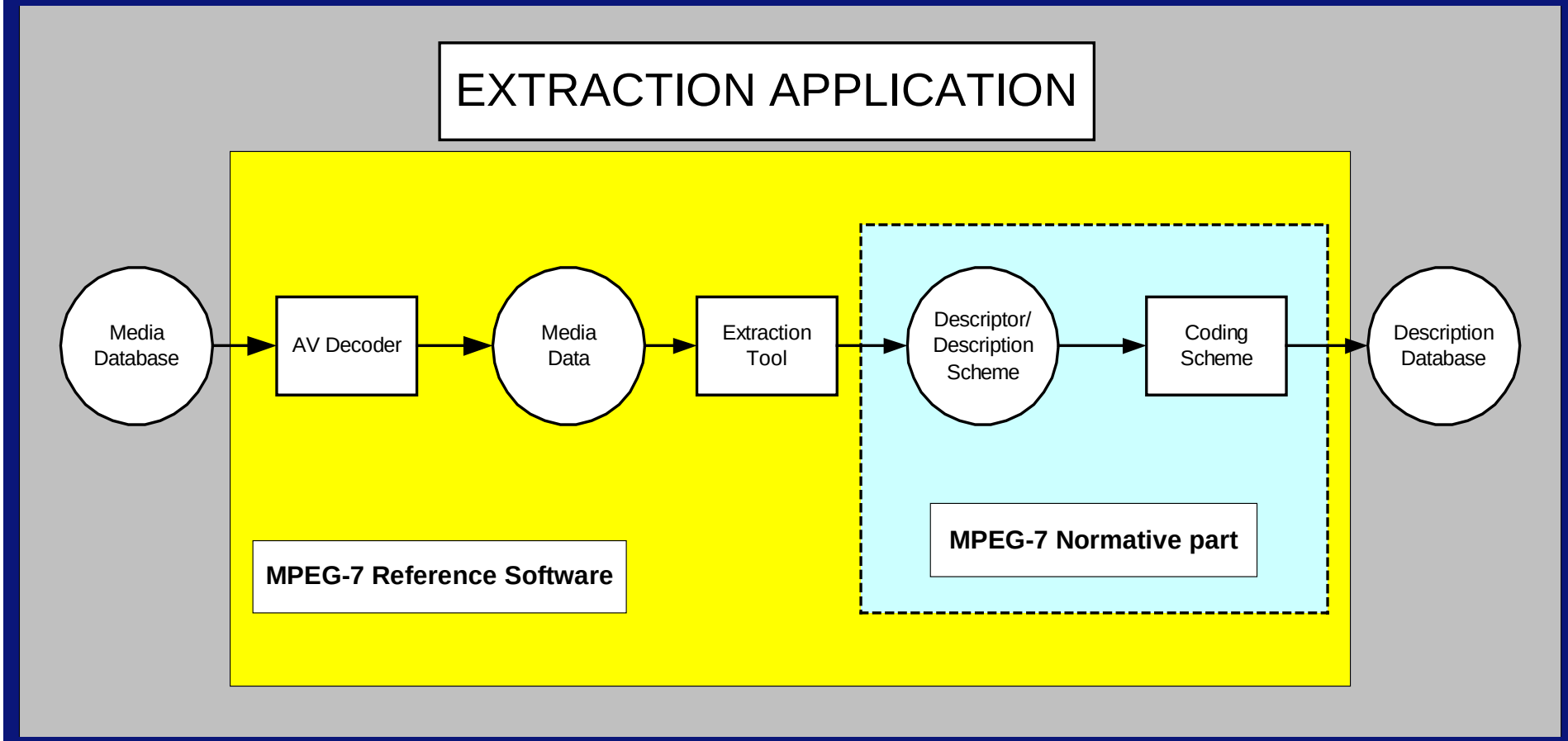
The following example gives an excerpt from Library of Congress Thesaurus of Graphical Material-I:

```
<Mpeg7>
  <Term termID="1">
    <Name>newsection 1</Name>
      <!-- newsection: 1 -->
      <Term termID="1970s">
        <Name>1970s</Name>
        <Term termID="Nineteen seventies" relation="US">
          <Name>Nineteen seventies</Name>
        </Term>
      </Term>
    </Term>
    <Term termID="3">
      <Name>newsection 3</Name>
      <!-- newsection: 3 -->
      <Term termID="3-legged racing">
        <Name>3-legged racing</Name>
        <Term termID="Three-legged racing" relation="US">
          <Name>Three-legged racing</Name>
        </Term>
      </Term>
    </Term>
    <!-- more terms -->
</Mpeg7>
```

MPEG-7: part 5 MDS conclusion

- MPEG-7 part 5 is the core part for all kind of descriptions
- interfaces to all other parts
- defines the overall structure of MPEG-7 descriptions
- is a good starting point even for simple descriptions without audio or video features

MPEG-7: part 6 Ref.-SW

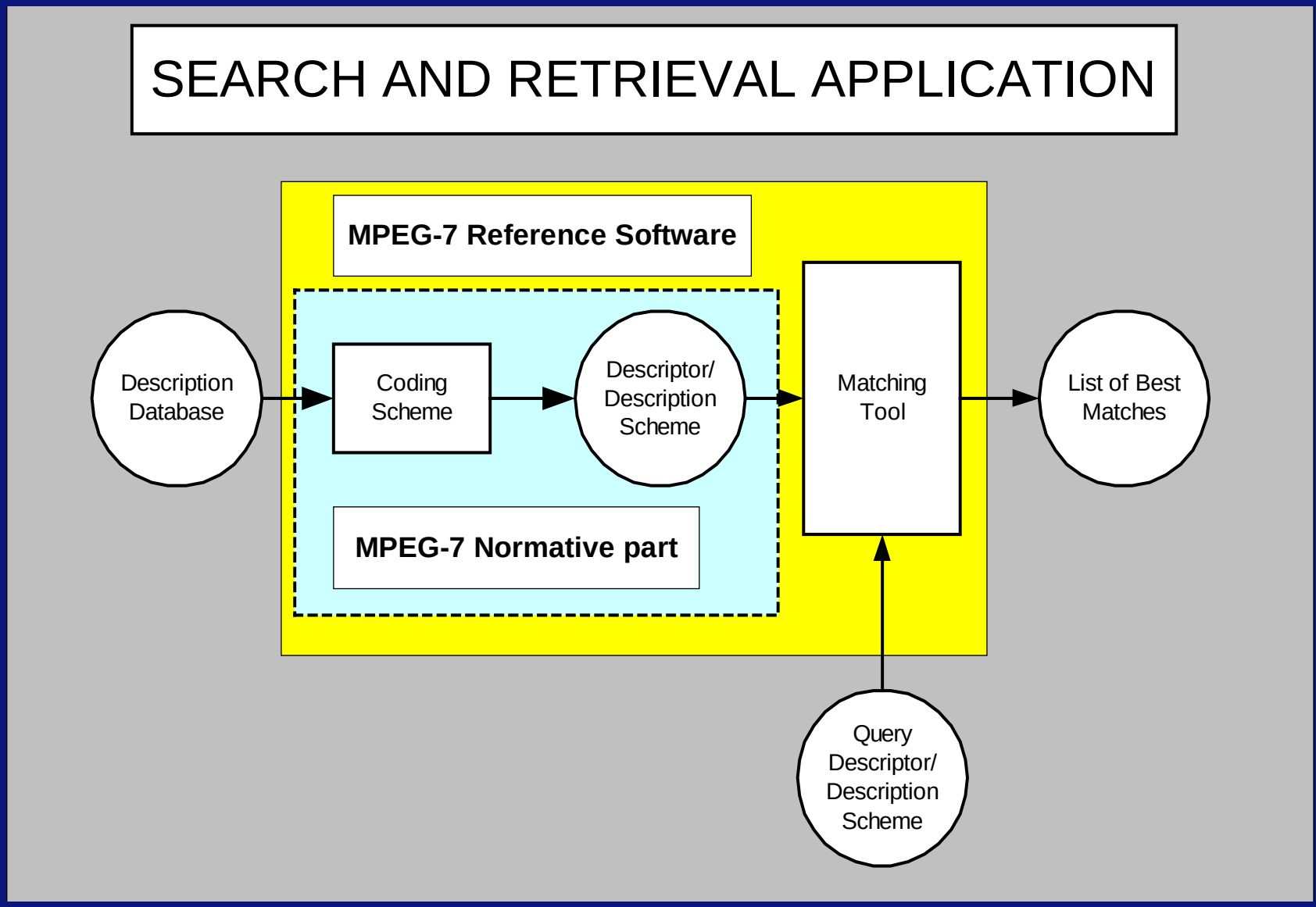


MPEG-7 Reference Software is hosted at TU-Muenchen

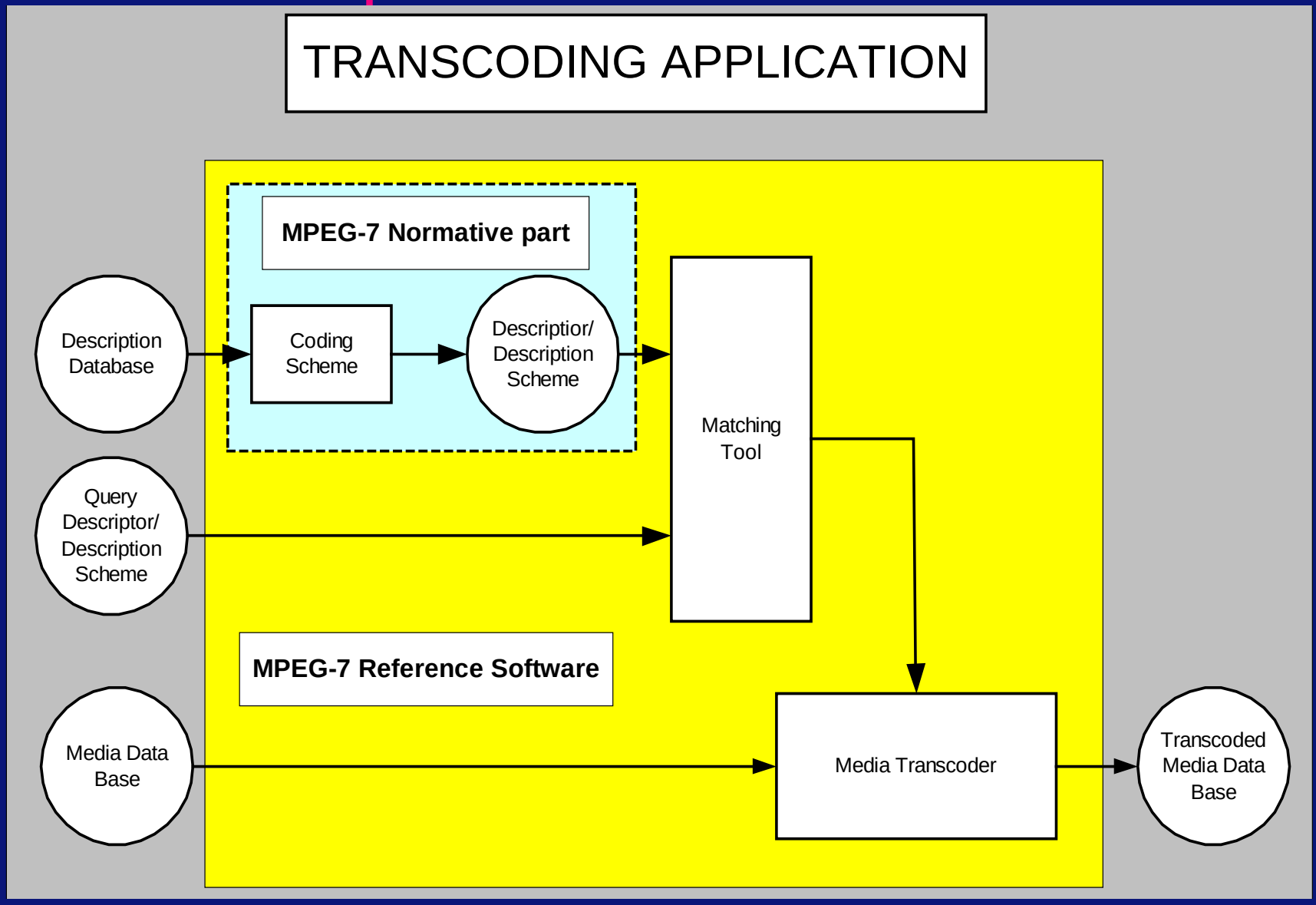
http://www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/mpeg7.html

===!"§==Systems=

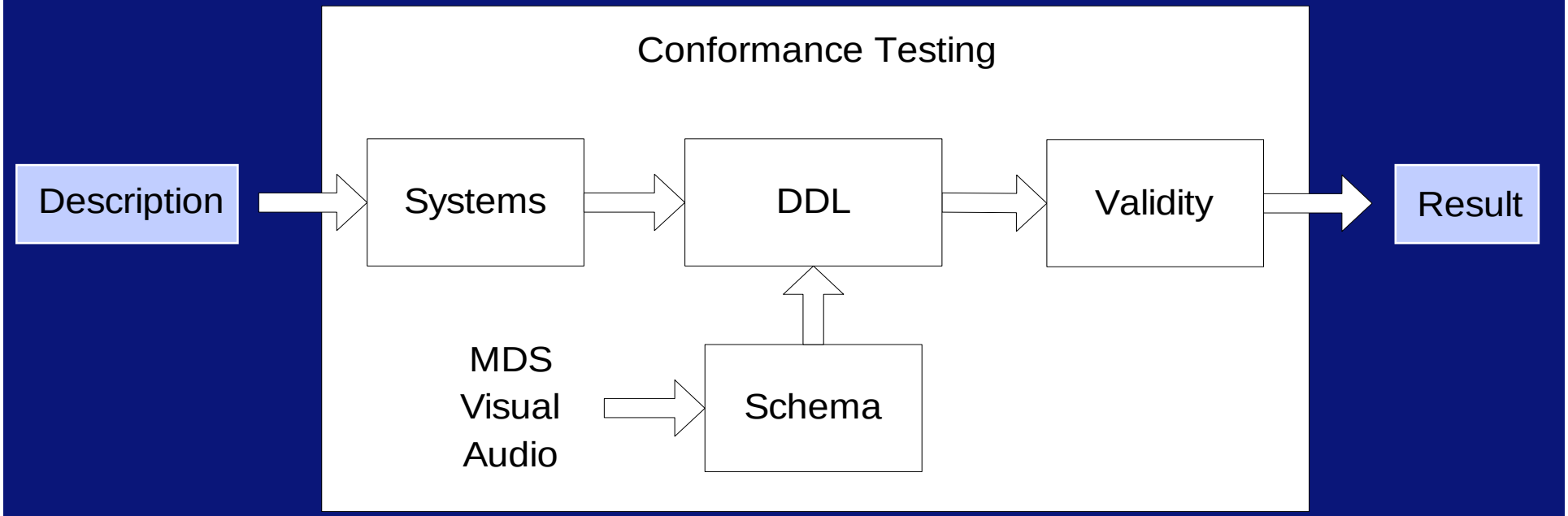
MPEG-7: part 6 Ref.-SW



MPEG-7: part 6 Ref.-SW



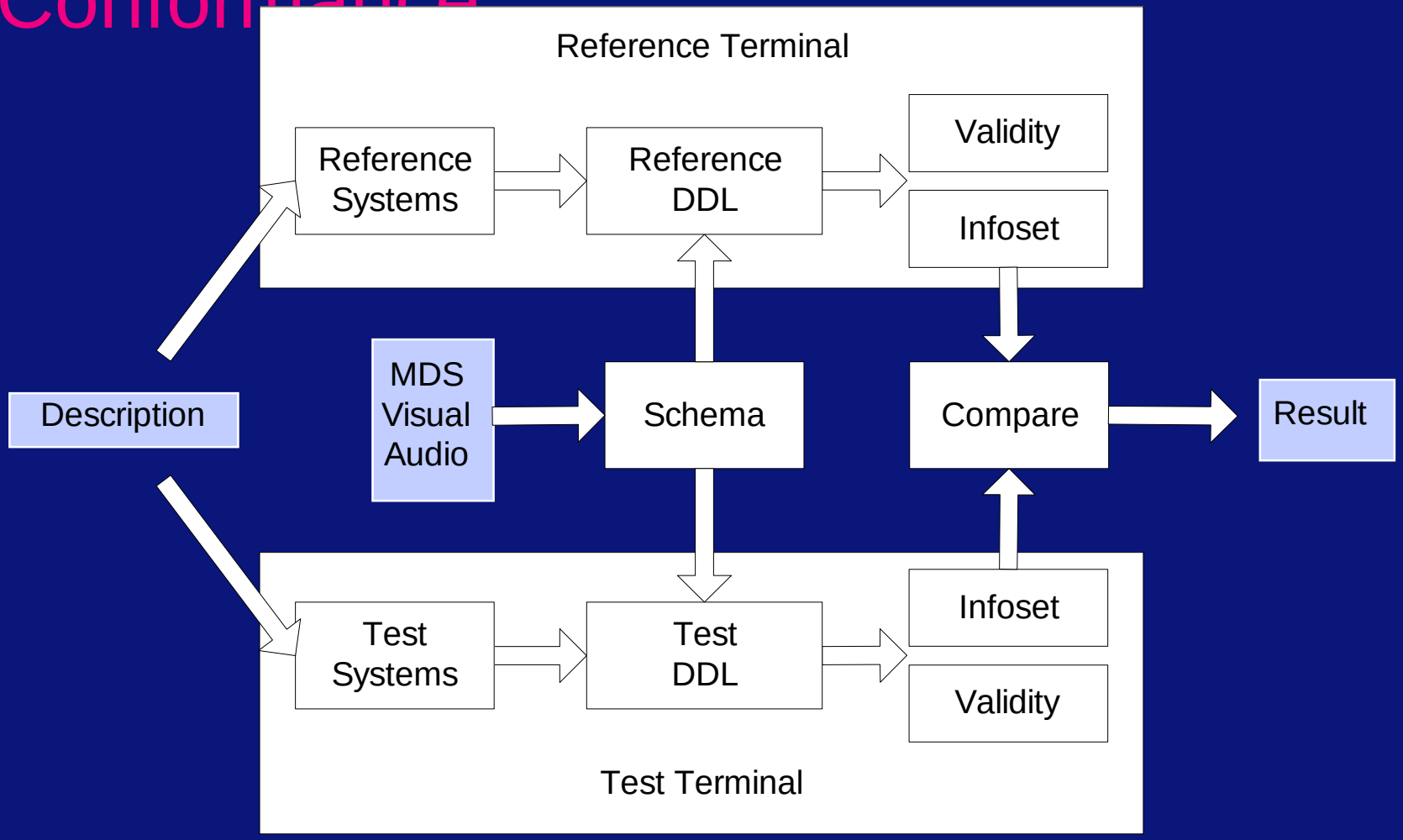
Structure of MPEG-7: part 7 Conformance



MPEG-7 Testing of Descriptions

===!"§==Systems=

Structure of MPEG-7: part 7 Conformance



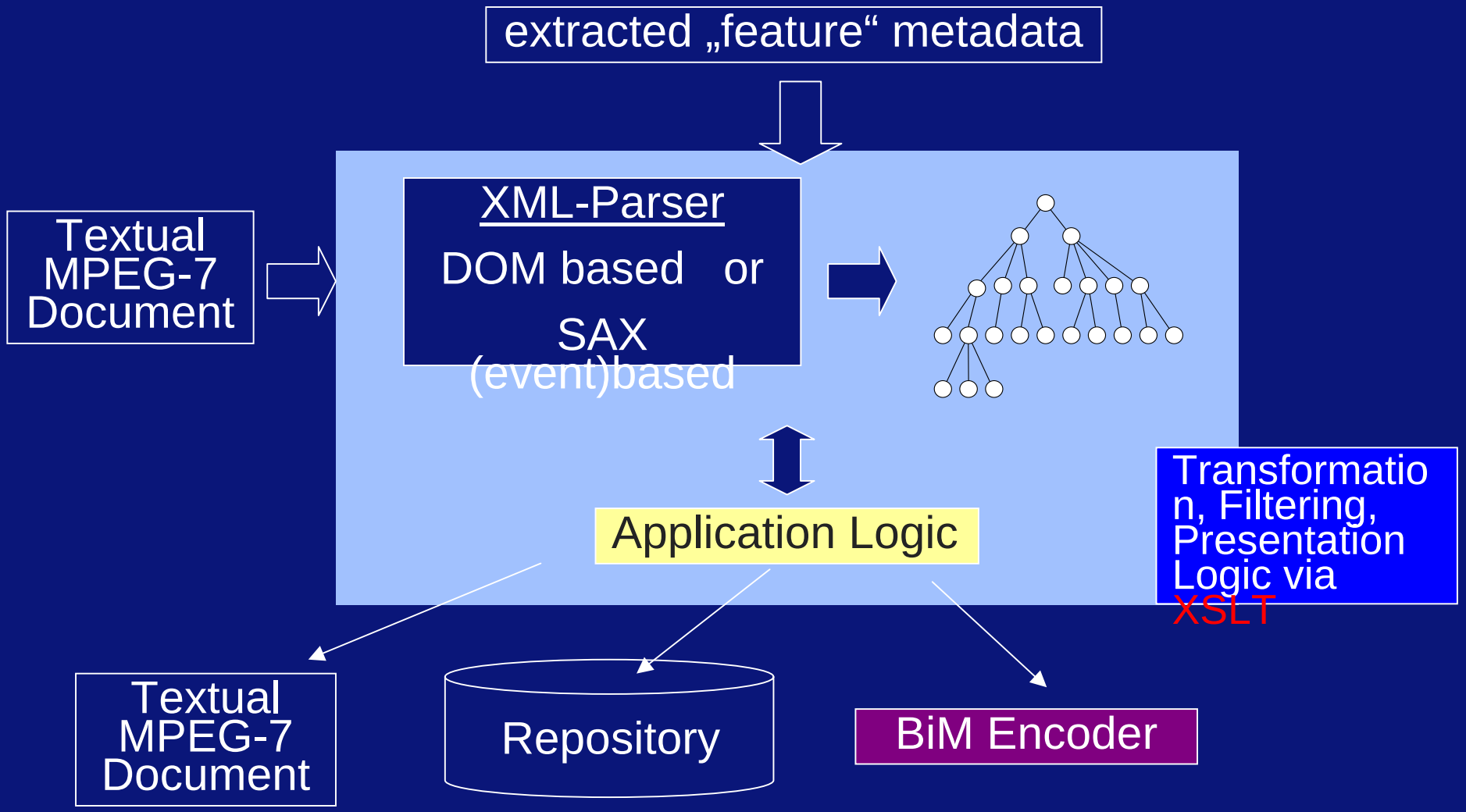
MPEG-7 Testing of Terminals

===!"§==Systems=

Structure of MPEG-7: part 8 Extraction & Use

- Provides further information on terminology, example instantiations
- Informative (nonnormative) methods for extraction
- Detailed examples for the core MDS-Part
- Detailed examples for Visual Part

MPEG-7: Extraction & Use



MPEG links and hints

Web-Pages:

MPEG-7 Alliance -----> http://www.mpeg-industry.com

- non-profit organisation
- applications and promotion to industry
- latest news abt. conferences and other events
- working groups / mailing lists

MPEG-Homepage -----> http://mpeg.telecomitalialab.com/

- overview documents
- relation to other MPEG standards

Acknowledgements

The presentation includes working material of:

- Bernhard Feiten (T-Systems Nova)
- A. B. Benitez (Columbia University)
- J.Herre (Frauenhofer IIS)
- M.Wollborn (Siemens)
- S.Quackenbush (IEEE)
- J. Martinez (TU Madrid)
- J. Smith (IBM)
- S.Herrmann (TU-München)
- C.Seyrat (Expway)

Thank you for your attention !