MPEG-7 as Metadata Framework for a Location Scouting System – An Evaluation Study

V. Gouaillier, L. Gagnon

S. Paquette, P. Poullaouec-Gonidec

Université de Montréal chaire UNESCO paysage et environnement

DC-2005, International Conference on Dublin Core and Metadata Applications
September 12 – 15, 2005

Outline

- Goals
- Database
- Metadata
- MPEG-7 Mapping
- System Architecture
- Conclusions

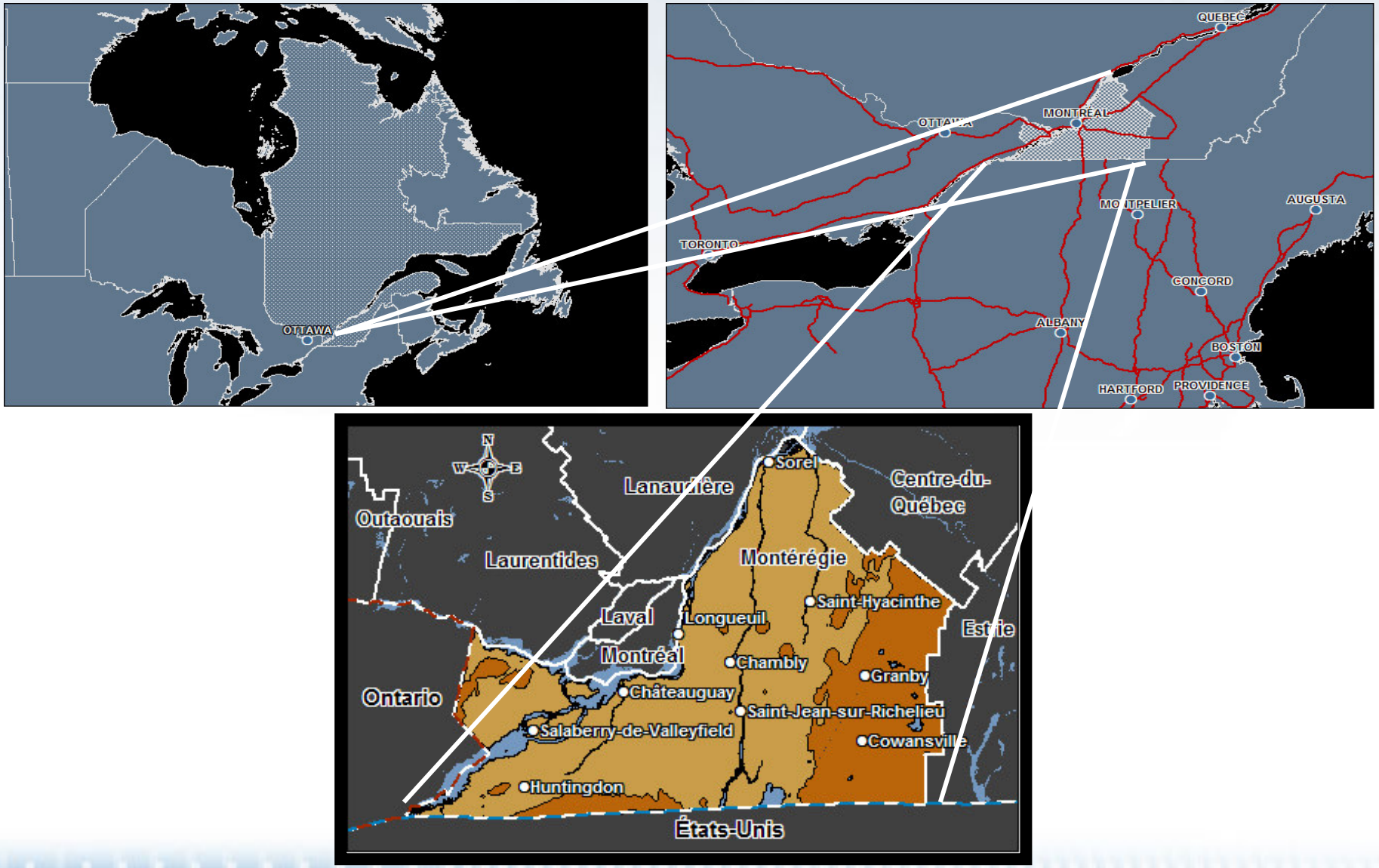
Project Goal

- Develop a software prototype for the **characterization and search of cinematographic sites** of the Monteregian region (Québec, Canada)
- Targeted applications
  - Film industry (location scouting)
  - Tourism industry
  - Promotion of the social, cultural and economical potential of the landscapes and infrastructures
  - Preservation and valorization of the patrimonial and environmental assets
  - Land planning

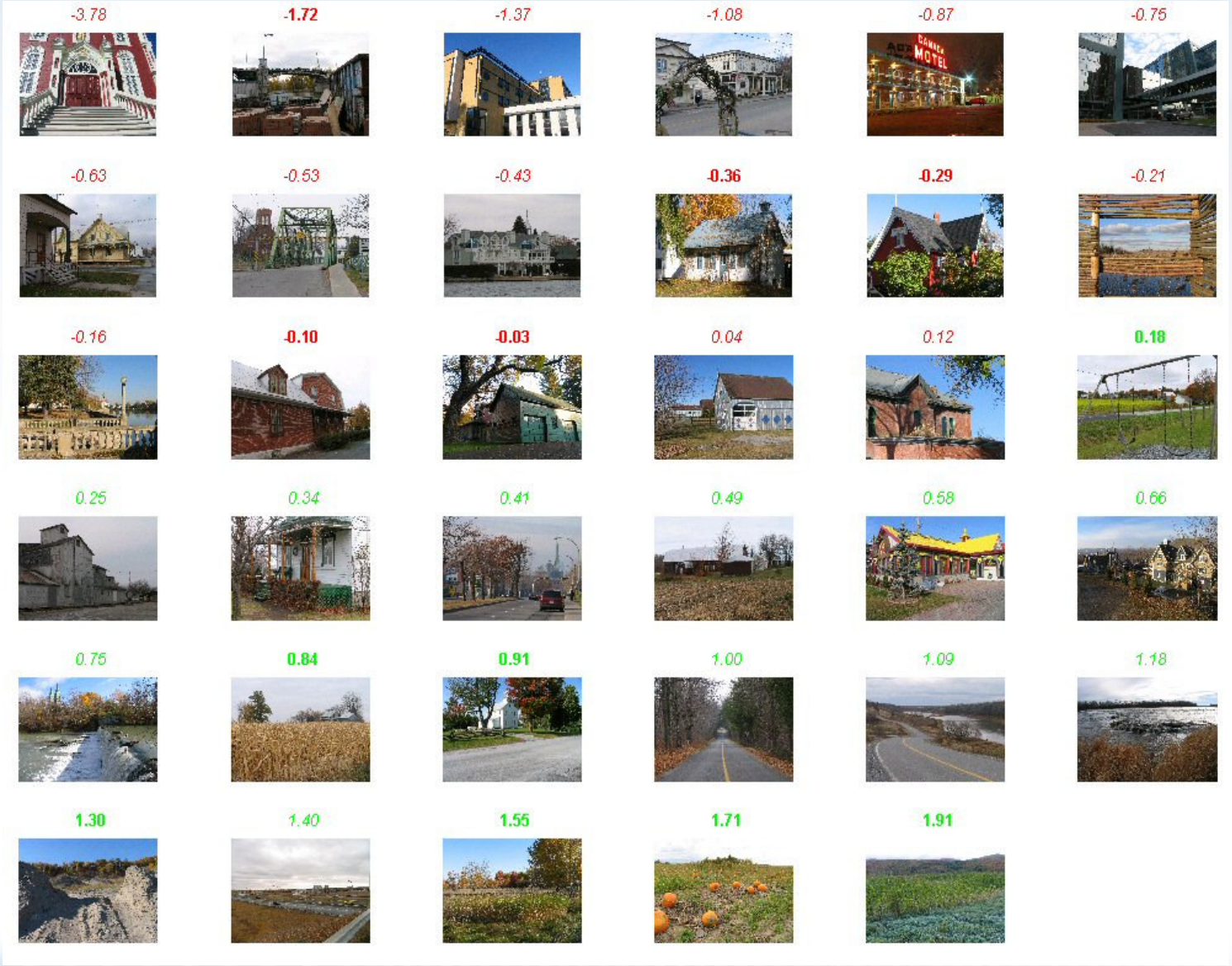
Technical Goals

- **Evaluate the use of MPEG-7 as metadata framework to describe the site (geographical) and associated image (visual content)**
- Offline indexing
  - Automatic for low-level visual content (image analysis tools)
  - Manual for geographical and landscape attributes
- Retrieval
  - Done through key words and/or search by similar images

The Monteregian Region (1)



The Monteregian Region (2)



Outline

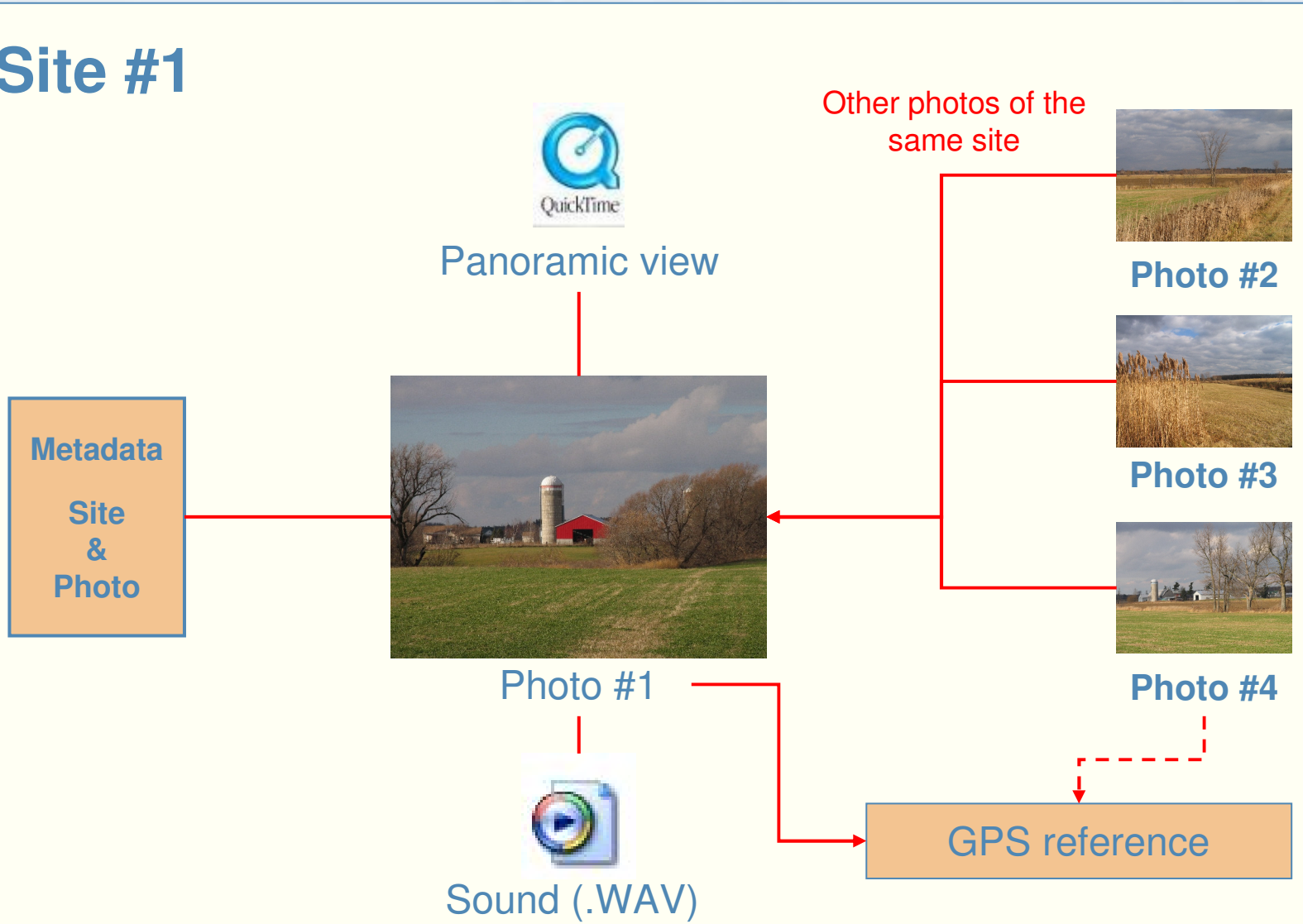
- Goals
- Database
- Metadata
- MPEG-7 Mapping
- System Architecture
- Conclusions

Database (1)

- Inventory of over 500 sites
- 1000 files of metadata (site and photo characteristics recorded from an expert-based evaluation)
- Approx. 3000 visual and audio elements
- GPS reference for cartographic queries

Database (2)

Site #1



Outline

- Goals
- Database
- Metadata
- MPEG-7 Mapping
- System Architecture
- Conclusions

Metadata

Site description

- **Geographic location** (geographic area, region, municipality, street, etc.)
- **Site characteristics** (public/private, function (residential, leisure, etc.))
- **Landscape type** (context (urban, rural, forest, etc.), scene characteristic scene (closed, open, etc.), site condition (habited, empty, abandoned, etc.))

Photo description

Structure	Semantics	Low-level visual features
• **Visual field attributes**	• **Biophysics attributes**	• **Color**
• **Foreground, middle ground, background**	• **Human settlements**	• **Texture**
• **Frame characteristics**	• **Agriculture**	• **Spectral shape**
• **Visual effects** (composition, kinetic effects, etc.)	• **Infrastructures**	• **Others**
	• **Aesthetic evocation** (realism, style, feeling evocation, etc.)	
	• **Contextual information**	

Site Description Example

```
<?xml version="1.0" encoding="utf-8" standalone="yes" ?>
- <Photo>
  + <Situation>
  - <Volet1>
    + <Lieu>
    - <Paysage>
      - <Type_paysage>
          <Forestier>Non</Forestier>
          <Agricole>Non</Agricole>
          <Routier>Non</Routier>
          <Ferroviaire>Non</Ferroviaire>
          <Littoral>Non</Littoral>
          <Villageois>Non</Villageois>
          <Urbain>Oui</Urbain>
          <Rocheux>Non</Rocheux>
          <Montagneux>Non</Montagneux>
          <Industriel>Non</Industriel>
          <Militaire>Non</Militaire>
          <Villegiature>Non</Villegiature>
          <Scene_de_rue>Non</Scene_de_rue>
          <Type_paysage_autres />
        </Type_paysage>
      - <Type_espace>
          <Contracte_renferme>Non</Contracte_renferme>
          <Dilate_elargi>Oui</Dilate_elargi>
          <Pictural>Non</Pictural>
          <Architectural>Oui</Architectural>
        </Type_espace>
      + <Etat_des_lieux>
      </Paysage>
    + <Milieu_Physique>
    + <Formes_etablissements_humains>
    + <Agriculture>
    + <Infrastructures>
    </Volet1>
  - <Volet2>
    + <Champ>
    + <Effets_visuels>
    - <Qualification_esthetique>
        <Realisme_scene_paysagere>Réaliste</Realisme_scene_paysagere>
      + <Style_esthetique>
      - <Ambiance_esthetique>
          <Poetique>Oui</Poetique>
          <Nostalgique_melancolique>Non</Nostalgique_melancolique>
          <Mysterieux>Non</Mysterieux>
          <Feerique>Non</Feerique>
          <Medieval>Non</Medieval>
          <Champetre>Non</Champetre>
          <Pastoral>Non</Pastoral>
          <Exotique>Non</Exotique>
          <Wilderness>Non</Wilderness>
          <Tenebreux>Non</Tenebreux>
          <Pittoresque>Non</Pittoresque>
          <Insolite>Non</Insolite>
          <Ludique>Non</Ludique>
          <Ambiance_esthetique_autres />
        </Ambiance_esthetique>
      + <Evocation_paysagere>
      </Qualification_esthetique>
    + <Qualification_contexte>
    </Volet2>
  </Photo>
```

Low-level Visual Descriptors Example

```
  </Descriptor>
- <Descriptor xsi:type="EdgeHistogramType">
    <BinCounts>0 2 0 4 1 2 4 4 6 2 0 4 6 4 2 0 3 4 6 2 0 5 3 3 2
      0 3 1 4 2 0 6 4 3 2 0 6 4 5 4 0 5 1 2 2 0 4 3 2 2 0 4 2 3 2
      1 7 2 2 1 0 4 2 0 4 0 4 0 3 2 0 2 1 0 3 0 2 0 0
      2</BinCounts>
  </Descriptor>
```

```
- <Descriptor xsi:type="ColorLayoutType">
    <YDCCoeff>18</YDCCoeff>
    <CbDCCoeff>32</CbDCCoeff>
    <CrDCCoeff>29</CrDCCoeff>
    <YACCoeff5>15 9 20 16 14</YACCoeff5>
    <CbACCoeff2>16 27</CbACCoeff2>
    <CrACCoeff2>17 8</CrACCoeff2>
  </Descriptor>
```

```
  </Descriptor>
- <Descriptor size="5" xsi:type="DominantColorType">
    <SpatialCoherency>0</SpatialCoherency>
  - <Values>
      <Percentage>12</Percentage>
      <ColorValueIndex>7 9 13</ColorValueIndex>
    </Values>
  - <Values>
      <Percentage>7</Percentage>
      <ColorValueIndex>21 18 12</ColorValueIndex>
    </Values>
  - <Values>
      <Percentage>7</Percentage>
      <ColorValueIndex>8 8 7</ColorValueIndex>
    </Values>
  - <Values>
      <Percentage>3</Percentage>
      <ColorValueIndex>14 15 16</ColorValueIndex>
    </Values>
  - <Values>
      <Percentage>1</Percentage>
      <ColorValueIndex>24 24 22</ColorValueIndex>
    </Values>
  </Descriptor>
```

Data Management Challenge

- Use a data management system based on an appropriate metadata structure
  - Taxonomy and vocabulary established by landscape and environment experts (about 400 words)
  - Data diversity (text and abstract mathematical attributes)
  - Search complexity (search by image content)

Outline

- Goals
- Database
- Metadata
- MPEG-7 Mapping
- System Architecture
- Conclusions

Why choosing MPEG-7?

- Need to describe **audio-visual documents**
- Need **structural and semantic information** description
- Offers a **set of low-level visual descriptors** for content-based image retrieval (from similar images or others)
- **Modular and flexible** (MPEG-7 allows to builds rich descriptors from predefined tools)
- XML-based : **human readability**, platform independance.
- Has tools to control vocabularies
  - allows to restrict the terms usage to a predefined list in a ClassicationScheme
  - map the landscape expert taxonomy

MPEG-7 Overview (1)

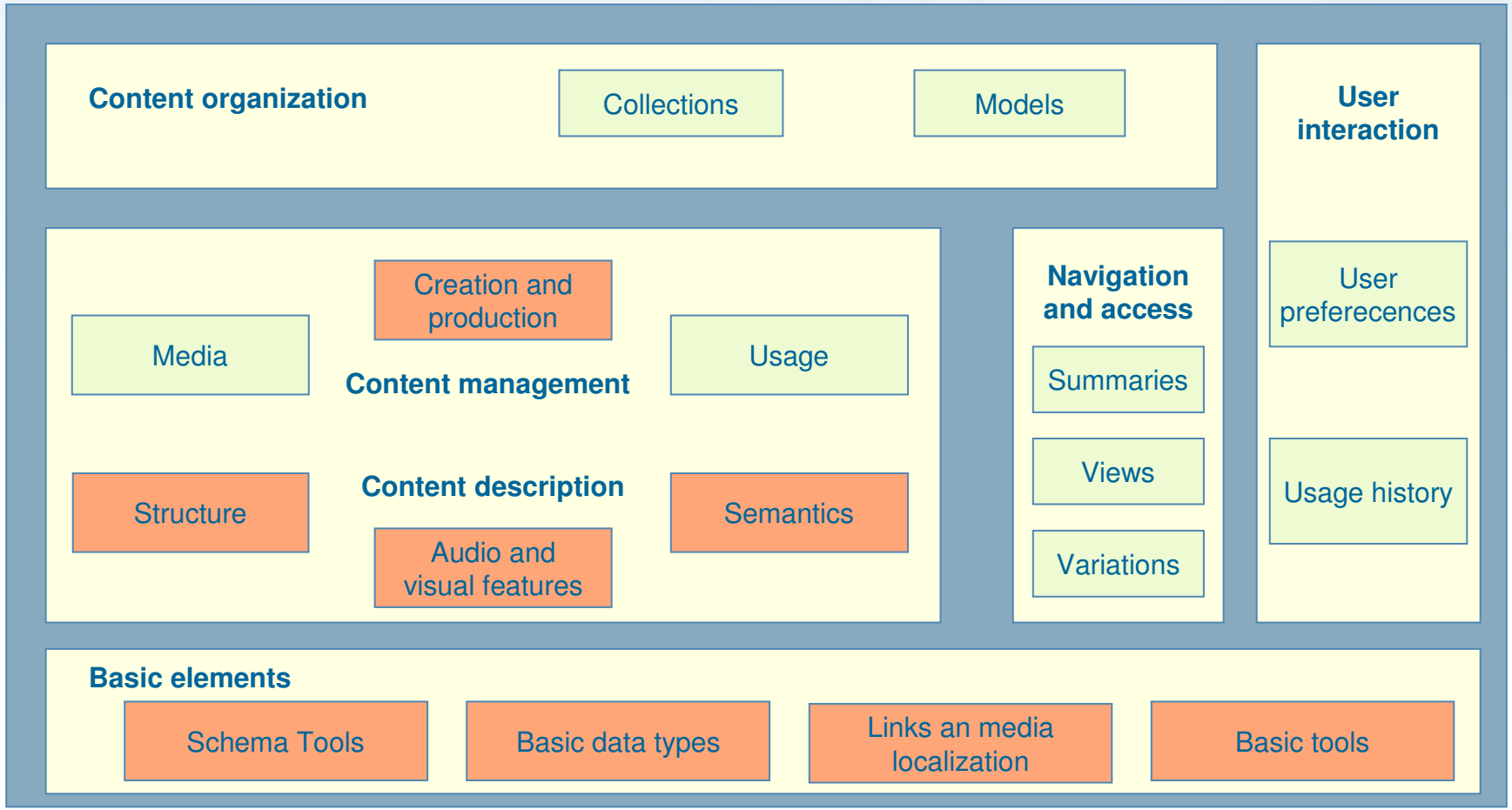
- Open standards for describing multimedia content to increase usability/accessibility
- Sept. 2001 (ISO/IEC 15938)
- Enables interoperability between systems which manages multimedia content
- Complements with MPEG-4 for solutions that require efficient streaming of content, content manipulation, indexing/retrieval

MPEG-7 Overview (2)

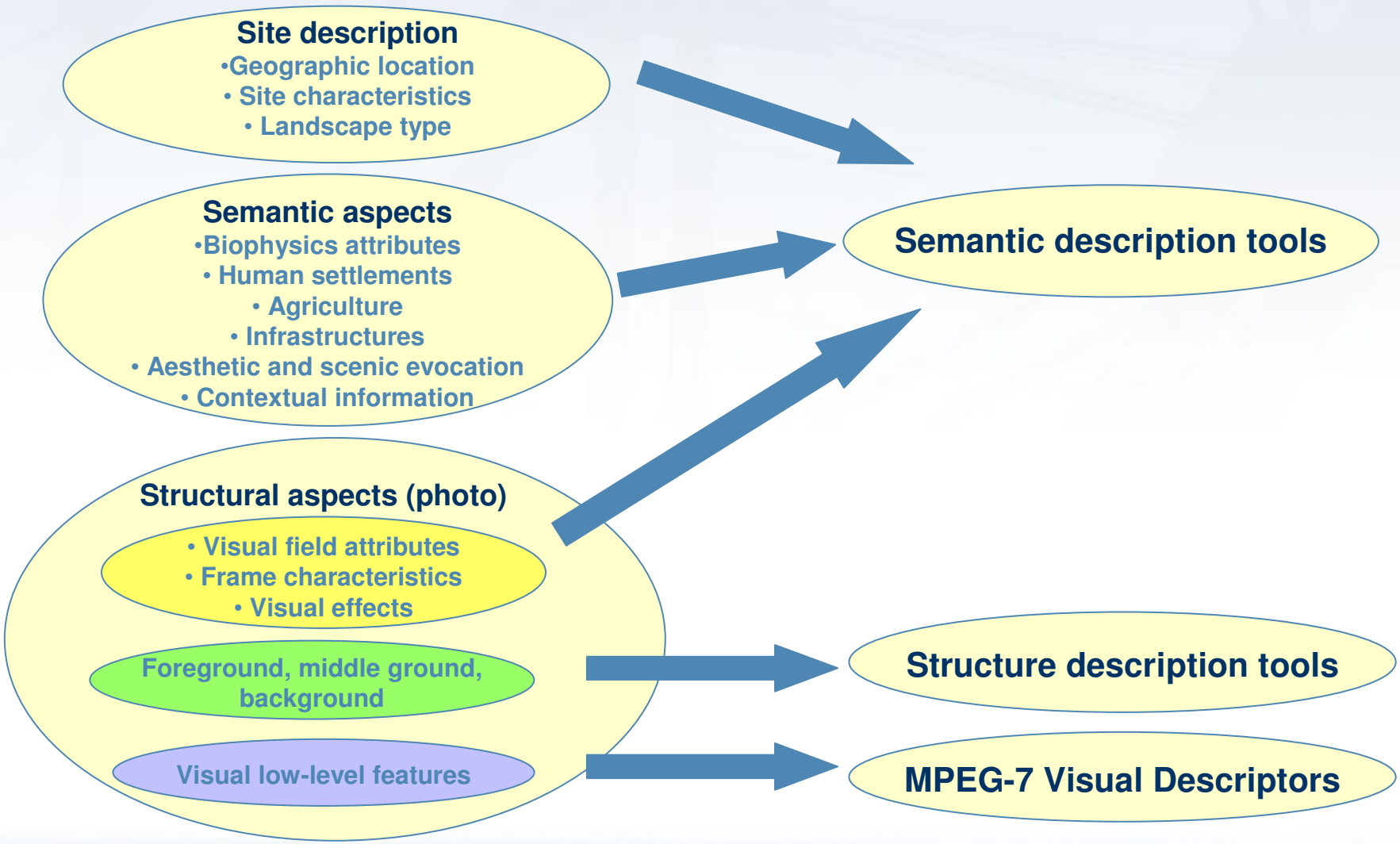
- Major contributors in multimedia
  - IBM, Canon, Hitachi, Ricoh, Ericsson, JVC, Philips, Siemens,...
- Harmonized with other standards
  - XML, XML Schema, URL, Dublin Core, MDD, ...
- Description encoding at different levels
  - catalogue (e.g. title, author)
  - semantic (e.g. who, what, when, where)
  - structural (e.g. spatio-temporal region, histogram, timbre, texture)

MPEG-7 Tools

Implemented tools in SIMIPE-Ciné:



Mapping to MPEG-7 (1)



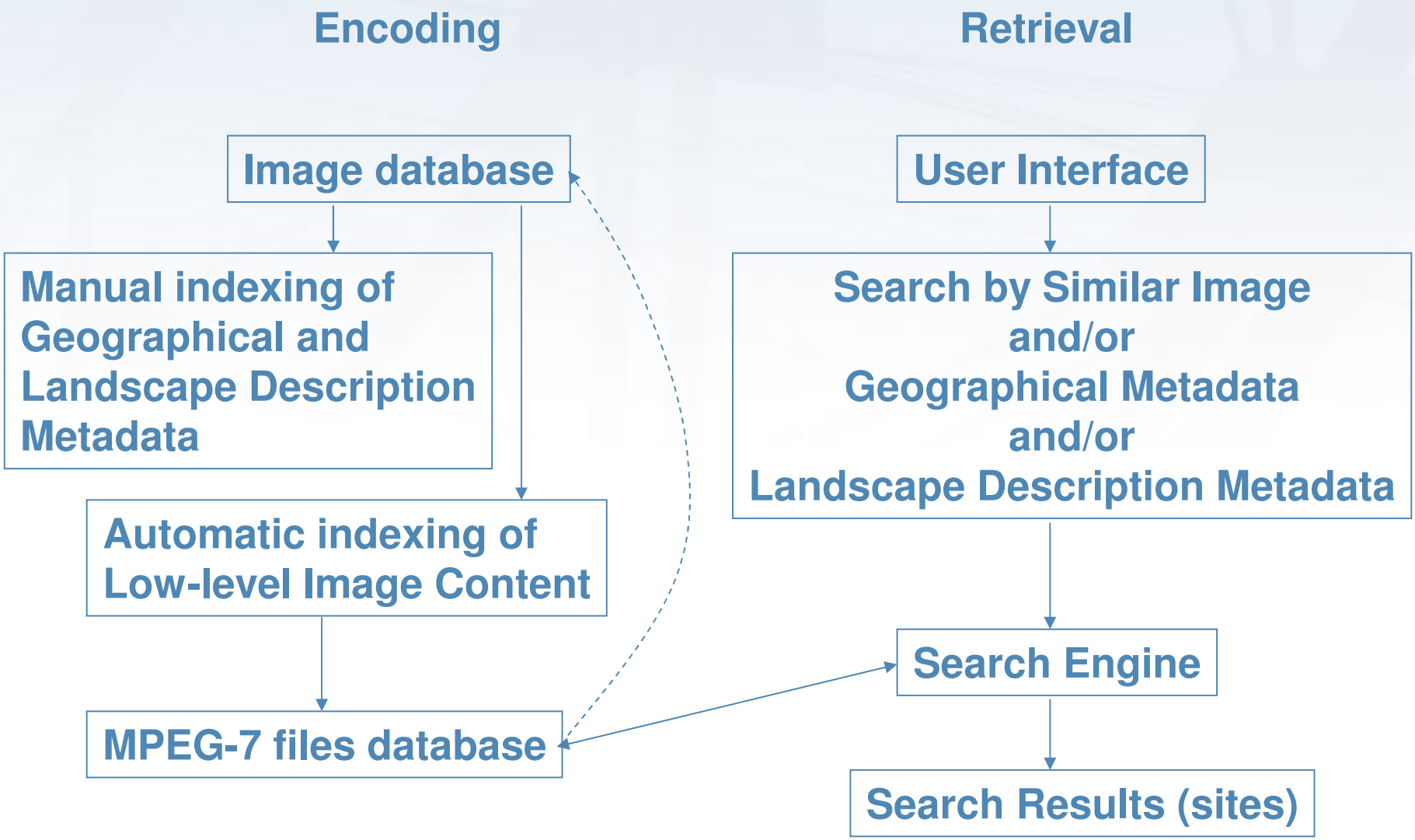
Mapping to MPEG-7 (2)

- MPEG-7 does not include predefined tools for the describtion of photographic structure metadata (visual FOV, scene framing, etc.)
  - Used the MPEG-7 semantic tools (SemanticType, SemanticStateType, Property)
- Only the presence or absence of scene plans can be mapped onto the MPEG-7 structure tools
  - StillRegion type allows to describe an image and its decomposition in plans using the SpatialDecomposition tag.
- Low-level visual description done using the standard MPEG-7 visual descriptors and others developed by CRIM (SpectrumShape).

Outline

- Goals
- Database
- Metadata
- MPEG-7 Mapping
- System Architecture
- Conclusions

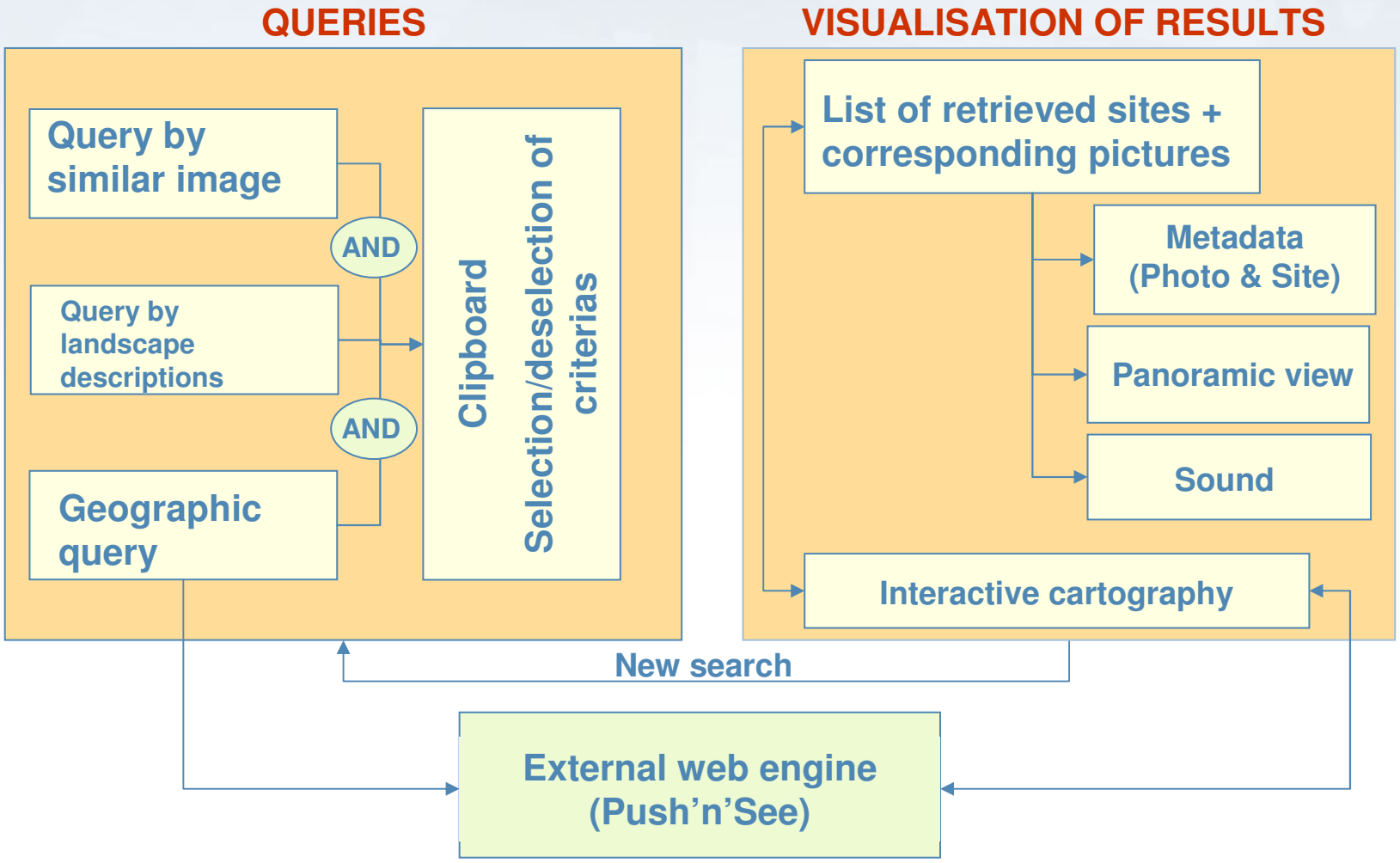
System Architecture



Queries Criterias

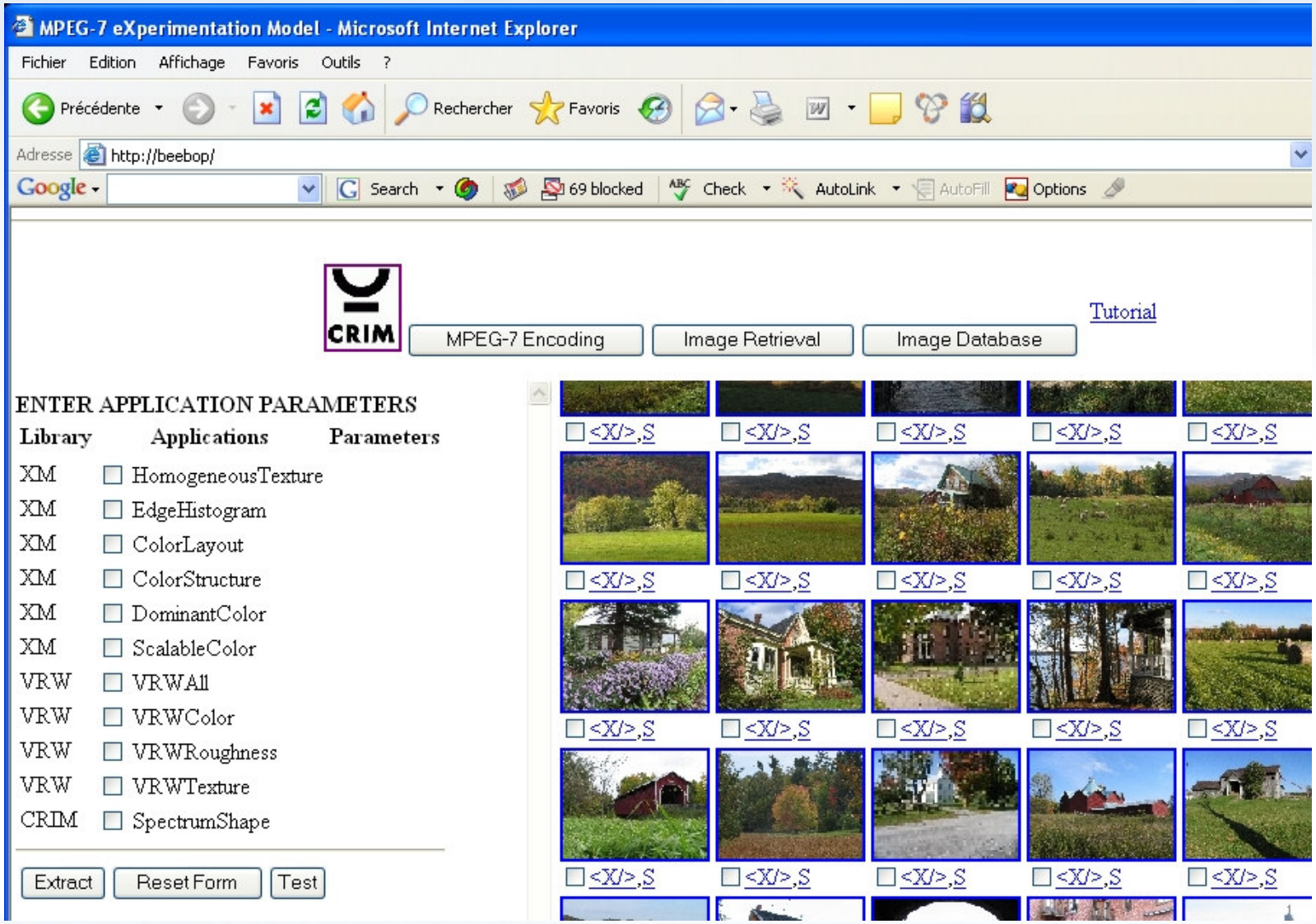
- Territorial, geophysical and infrastructural aspects of the sites
- Context (ambiant sound, climatic and environmental conditions)
- Aesthetic assessment of landscape scene
- Structure and characteristics of related photos
- Visual similarity of sites
- Geographic attributes

User Interface Functionalities



Mock-up Snapshot

Search by Visual Content



Similarity Search Example

Query

- Edge Histogram
- Color Layout
- Color Structure
- Dominant Color
- Scalable Color

Retrieval

Most similar

Most dissimilar

Conclusions

- MPEG-7 strongs for audio-visual content description but
  - **Less adapted to geographic site description**
  - « Heavy-weight » approach (**long descriptions**)
  - No efficient way for **structural description of some photo characteristics** (i.e. visual field attributes, frame characteristics, visual effects)
- Despite of that, representation of both structural and semantic information relative to sites as been achieved
- Resulting metadata structure respects landscape and environment description taxonomy
- However, no strong reason to use MPEG-7 if search by audio-visual content not part of the specifications