A Common Infrastructure for Cultural Heritage Applications

L. Van Eycken, B. DeKnuydt, L. Van Gool
K.U. Leuven
Requirements for a Cultural Heritage infrastructure

The concepts behind a common infrastructure

Cultural heritage data

Cultural heritage information management

Suggested standards and formats

Implementation examples
Requirements for a CH infrastructure

- Fit small as well as large applications
- Fit a diversity of applications
  - Show/navigate the data
  - Present the data in an exhibition
  - Preserve the data (need for sustainability)
- Fit a wide range of sources (CH = multidisciplinary) ⇒ not tailored to one discipline
- Data is dynamic (created, augmented & modified)
- As technology independent as possible
Requirements for a Cultural Heritage infrastructure

The concepts behind a common infrastructure

Cultural heritage data

Cultural heritage information management

Suggested standards and formats

Implementation examples
Concepts behind a common infrastructure

Cultural Heritage applications

- Simple applications: a pipeline
  - Input
    - Scan object
  - Tool 1
    - Clean 3D
  - Tool 2
    - Annotate
  - Output
    - Display 3D & annotations

- More complex applications
  - Different sub-applications
    - generated/used at different times and places
    - exchange CH data
  - Each sub-application = pipeline
  - Data is combined and iterated upon
Concepts behind a common infrastructure

A typical cultural heritage pipeline (by Nick Ryan)

1. Multi-lingual and semantic processing
2. Databases and technology management
3. Mobile, wearable, ambient systems

Project design, background research → Field survey and/or excavation → Interpretation and analysis → Scholarly publication and archiving → Public presentation


Loops and iterations not shown!

Tools are different for different application domains (sites, museums, tourism, education, ...)!
The information processing view

- Data synthesis & modification tools
- Information store
- Data input tools
- Data output tools

- Real applications: only a subset of tools & stored data, but all stores should be based on the same foundations
- Store does not necessarily correspond to physical storage
Requirements for a Cultural Heritage infrastructure

The concepts behind a common infrastructure

Cultural heritage data

Cultural heritage information management

Suggested standards and formats

Implementation examples
The diversity of cultural heritage data

- **Typical usage**

<table>
<thead>
<tr>
<th></th>
<th>Preservation</th>
<th>Exchange</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sustainability</td>
<td>important</td>
<td>less relevant</td>
<td>less relevant</td>
</tr>
<tr>
<td>Standards used</td>
<td>few non-proprietary</td>
<td>few</td>
<td>application driven</td>
</tr>
<tr>
<td>Compactness</td>
<td>desirable</td>
<td>desirable</td>
<td>irrelevant</td>
</tr>
<tr>
<td>Fast running</td>
<td>irrelevant</td>
<td>less relevant</td>
<td>important</td>
</tr>
</tbody>
</table>

- avoid format conversions if possible

- **Type of data**
  - usually heterogeneous
  - relations important ⇒ rich linking facilities needed

- **Accuracy**
  - recorded: keep at maximum accuracy; lossless compression
  - illustrations: look/feel/sound good; lossy compression
The Cultural Heritage Data Object (CHDO)

CHDO = basic unit for storing CH data and its relations

- **Container format**
  - keep all relevant CH data linked together
  - store global info (e.g., IPR)

- **Basic preservation unit**
  - contains multiple data types, incl. semantic info
  - neutral, non-3D centric view

- **Useful as CH exchange unit**
  - usable with multiple standards or exchange formats

<table>
<thead>
<tr>
<th>Container</th>
<th>Possible standards:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illustration image</td>
<td>METS, MPEG-21, ...</td>
</tr>
<tr>
<td>Recorded image</td>
<td>JPEG, JPEG2000, ...</td>
</tr>
<tr>
<td>3D model</td>
<td>TIFF/LZW, PNG, ...</td>
</tr>
<tr>
<td>Annotation</td>
<td>Collada, X3D</td>
</tr>
<tr>
<td>CIDOC-CRM data</td>
<td>RDF, Dublin Core, ...</td>
</tr>
<tr>
<td>CIDOC XML</td>
<td></td>
</tr>
</tbody>
</table>
Standards for container classes

- Existing non-specific container standards
  - DIDL from MPEG-21 (multimedia framework)
  - XPackage & RDF (W3C standards)
  - METS (used in digital libraries)

- Current Epoch’s choice: METS
  - active development
  - more interaction with other standards

Note: Concept more important than standard
- Requirements for a Cultural Heritage infrastructure
- The concepts behind a common infrastructure
- Cultural heritage data
  - Cultural heritage information management
- Suggested standards and formats
- Implementation examples
Collection of CHDO

CHDO granularity depends on:
- data usage
- implementation restrictions

Semantic network as a single CHDO

... or a distributed semantic network
Implementation of the information store

= any digital library but ...

with a flexible digital library format which is easy to extend and adapt to specific Epoch needs

- Any implementation can be used
  - Simple one: e.g., set of files in agreed upon format
  - Application specific one: e.g., MAD semantic database
  - More general one: e.g. Fedora
    - data manipulation tools available
    - fine grained access control
    - large user community
- Requirements for a Cultural Heritage infrastructure
- The concepts behind a common infrastructure
- Cultural heritage data
- Cultural heritage information management
- Suggested standards and formats
- Implementation examples
Suggested standards and formats

No new standards: enough available for our goals

- Guidelines for exchange and preservation of sound & images (audio, vector graphics, 2D images, video)
  - see Epoch web site

- Proposal for exchange of 3D
  - Collada for data (Epoch uses “Collada light”)
  - U3D for 3D data compression

Note: These guidelines will evolve over time, also for preservation formats
Requirements for a Cultural Heritage infrastructure

The concepts behind a common infrastructure

Cultural heritage data

Cultural heritage information management

Suggested standards and formats

Implementation examples
Implementation examples

Tool chains

- Purpose
  - testing tool interoperability
  - gaining more experience in exchange problems and formats

- Fedora as content management system in the tool chains:
  - A virtual visitor centre (also multimedia dissemination)
  - Archaeological field survey
  - Semantic browsing of 3D artefacts
Epoch provides some guidelines

- for data exchange formats
- to select a proper content management system, such as Fedora

These guidelines

- allow application specific optimal solutions
- anticipate new technologies and solutions