Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

CASPAR: Components for a Science Data Infrastructure – preservation and re-use of data

David Giaretta

Alliance for Permanent Access

CCSDS.ORG
The Consultative Committee for Space Data Systems

DCC

PARSE insight

Overview

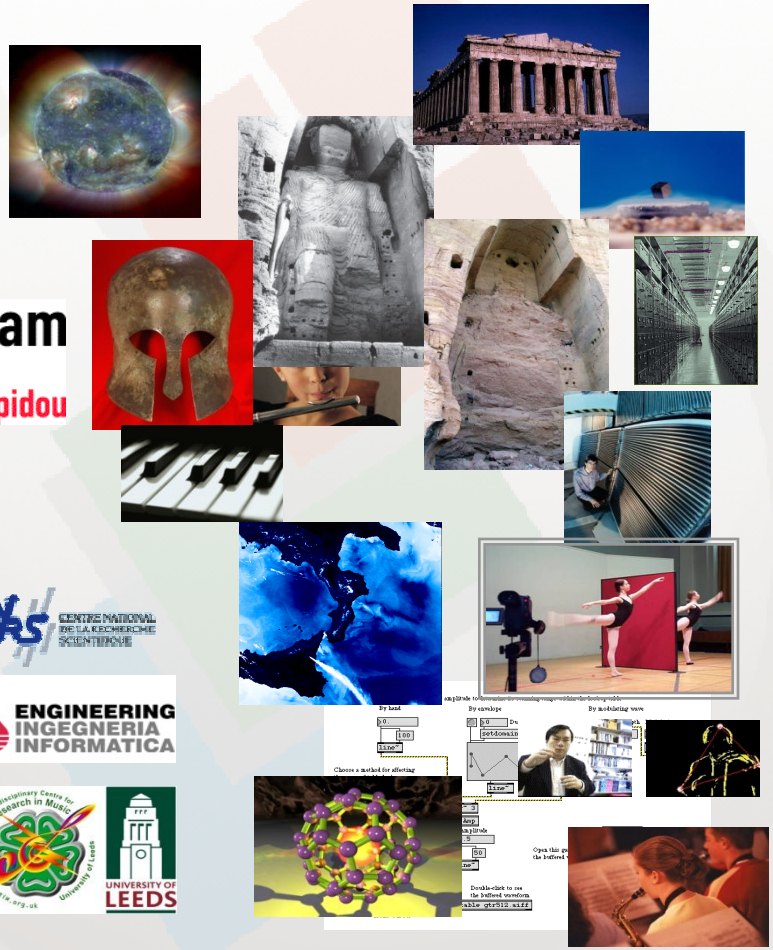
- CASPAR
- OAIS
- Threats and Solutions
- Validation

Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

CASPAR Project

EU FP6 Integrated Project

Total spend approx. 16MEuro (8.8 MEuro from EU)



http://www.casparpreserves.eu

Digital Preservation

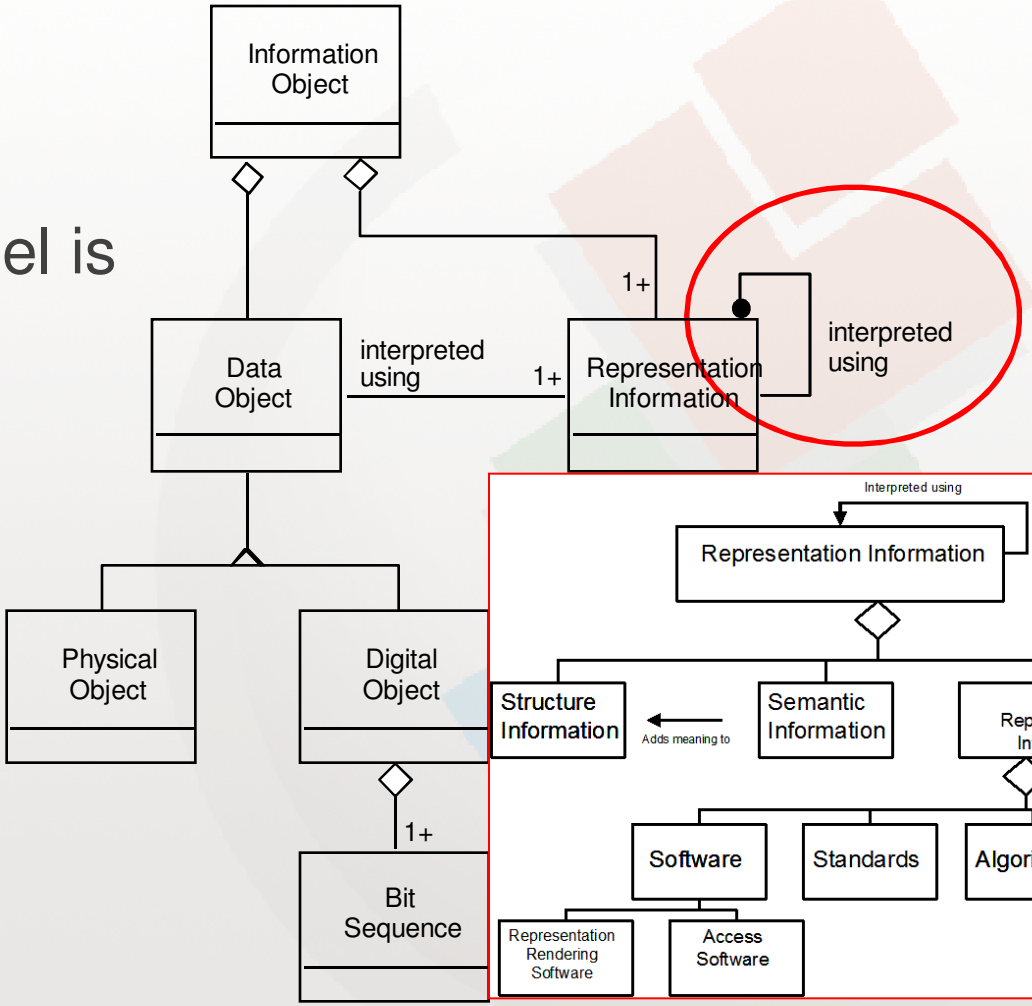
- Ensure that digitally encoded information are understandable and usable over the long term
  - Long term could start at just a few years
- Easy to make claims
  - Difficult to provide proof
- Reference Model for Open Archival Information System (ISO 14721)
  - The basic standard for work in digital pres.
  - Defines terminology and compliance criteria

Information Model & Representation Information

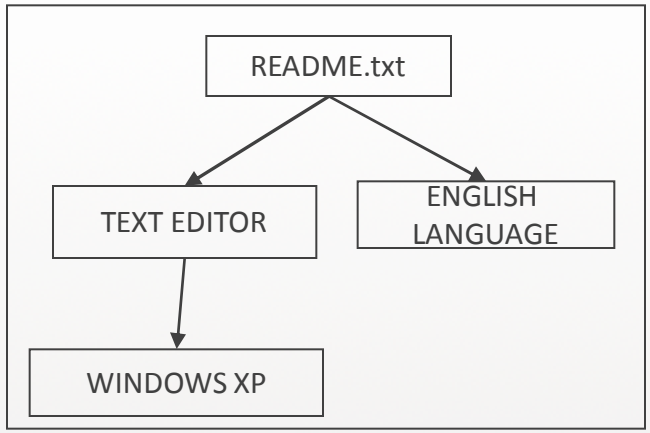
The Information Model is key

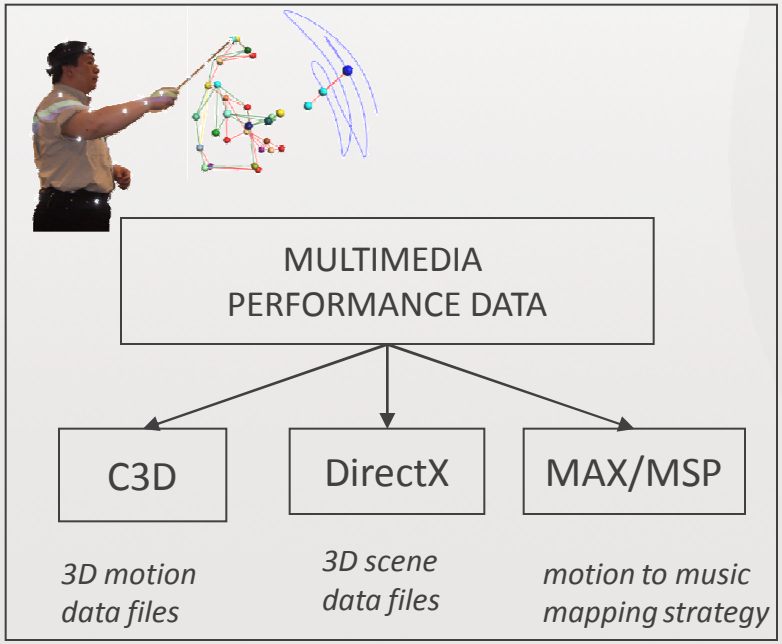
Recursion ends at KNOWLEDGEBASE of the DESIGNATED COMMUNITY

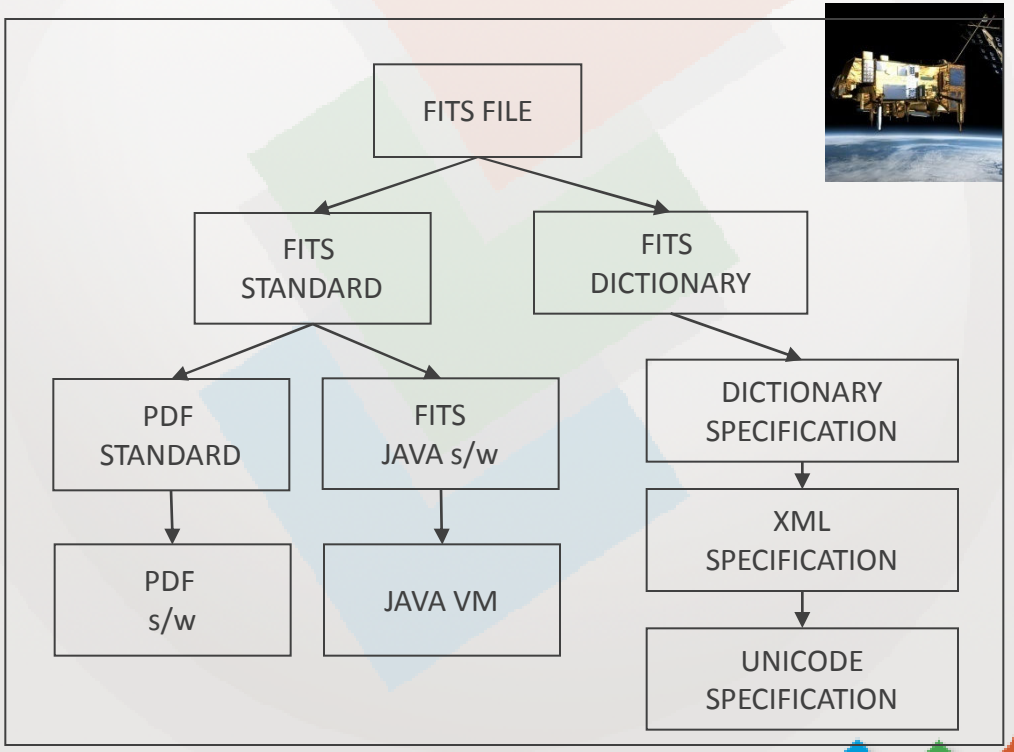
(this knowledge will change over time and region)

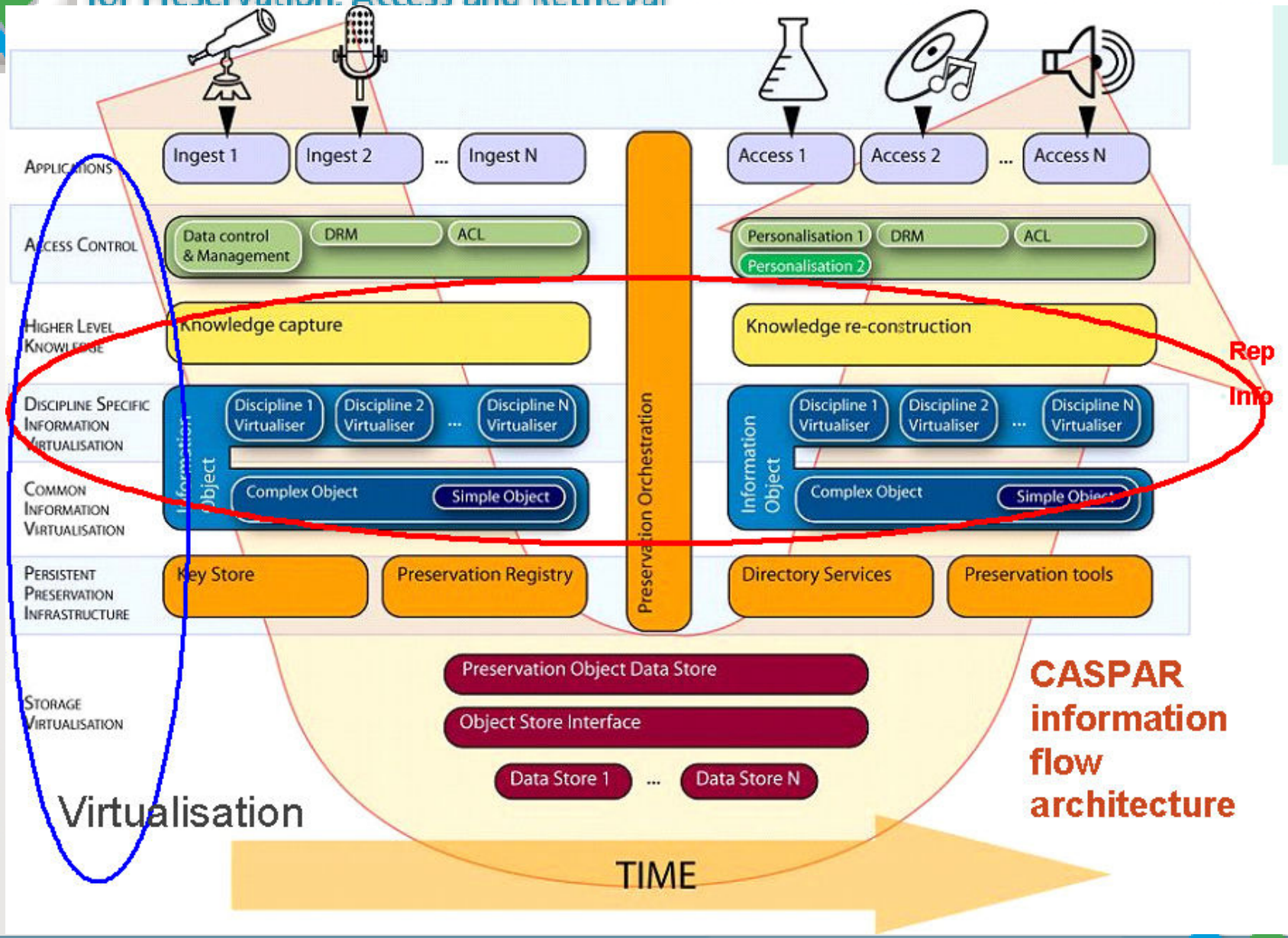


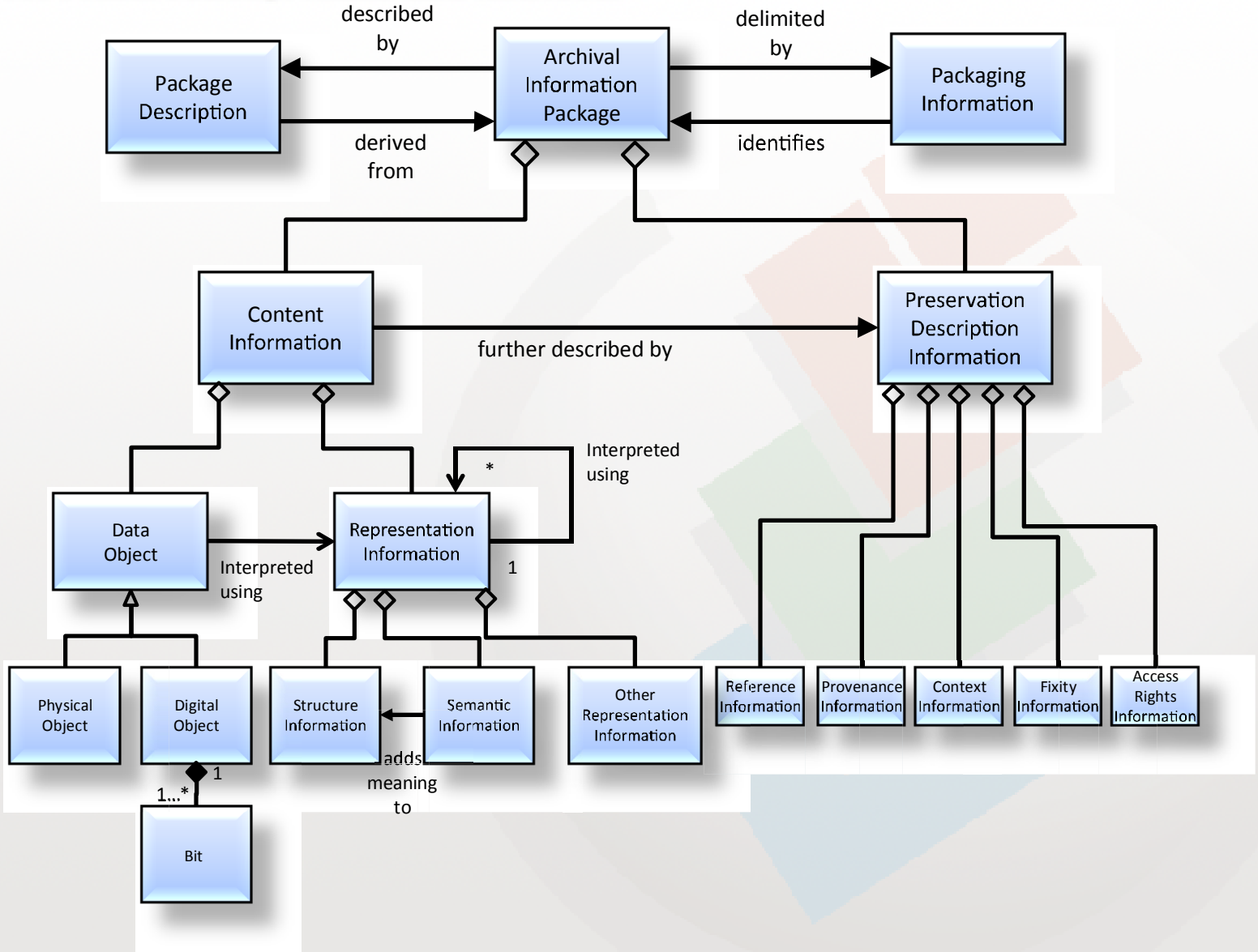
Modules and Dependencies: defining the Designated Community



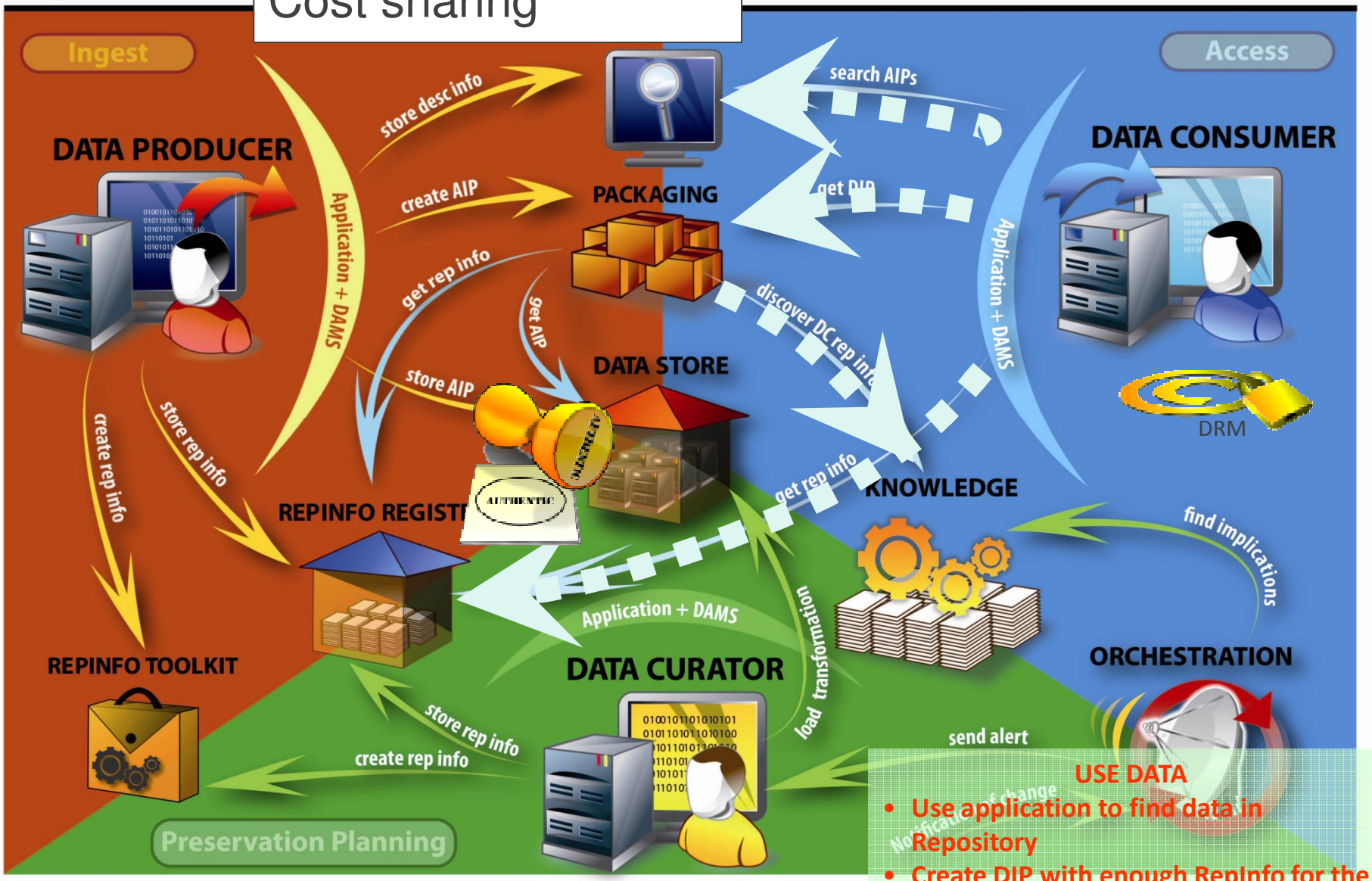





Ingest 1
Ingest 2
...
Ingest N
Access 1
Access 2
...
Access N
Applications
Access Control
Data control & Management
DRM
ACL
Personalisation 1
DRM
ACL
Personalisation 2
Higher Level Knowledge
Knowledge capture
Knowledge re-construction
Discipline Specific Information Virtualisation
Discipline 1 Virtualiser
Discipline 2 Virtualiser
...
Discipline N Virtualiser
Information Object
Common Information Virtualisation
Complex Object
Simple Object
Persistent Preservation Infrastructure
Key Store
Preservation Registry
Preservation Orchestration
Directory Services
Preservation tools
Storage Virtualisation
Preservation Object Data Store
Object Store Interface
Data Store 1
...
Data Store N
Rep
Info
Virtualisation
TIME
CASPAR information flow architecture


described by
delimited by
Package Description
Archival Information Package
Packaging Information
derived from
identifies
Content Information
further described by
Preservation Description Information
Interpreted using
*
Data Object
Representation Information
Interpreted using
1
Physical Object
Digital Object
Structure Information
Semantic Information
Other Representation Information
Reference Information
Provenance Information
Context Information
Fixity Information
Access Rights Information
adds meaning to
1
1...*
Bit

Cost sharing



Preservable infrastructure

USE DATA

- Use application to find data in Repository
- Create DIP with enough RepInfo for the user (via DC profile)
- Obtain more RepInfo from Registry if necessary

Threat	Requirement for solution
Users may be unable to understand or use the data e.g. the semantics, format, processes or algorithms involved	
Non-maintainability of essential hardware, software or support environment may make the information inaccessible	
The chain of evidence may be lost and there may be lack of certainty of provenance or authenticity	
Access and use restrictions may make it difficult to reuse data, or alternatively may not be respected in future	
Loss of ability to identify the location of data	
The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future	
The ones we trust to look after the digital holdings may let us down	

Accelerated Lifetime tests

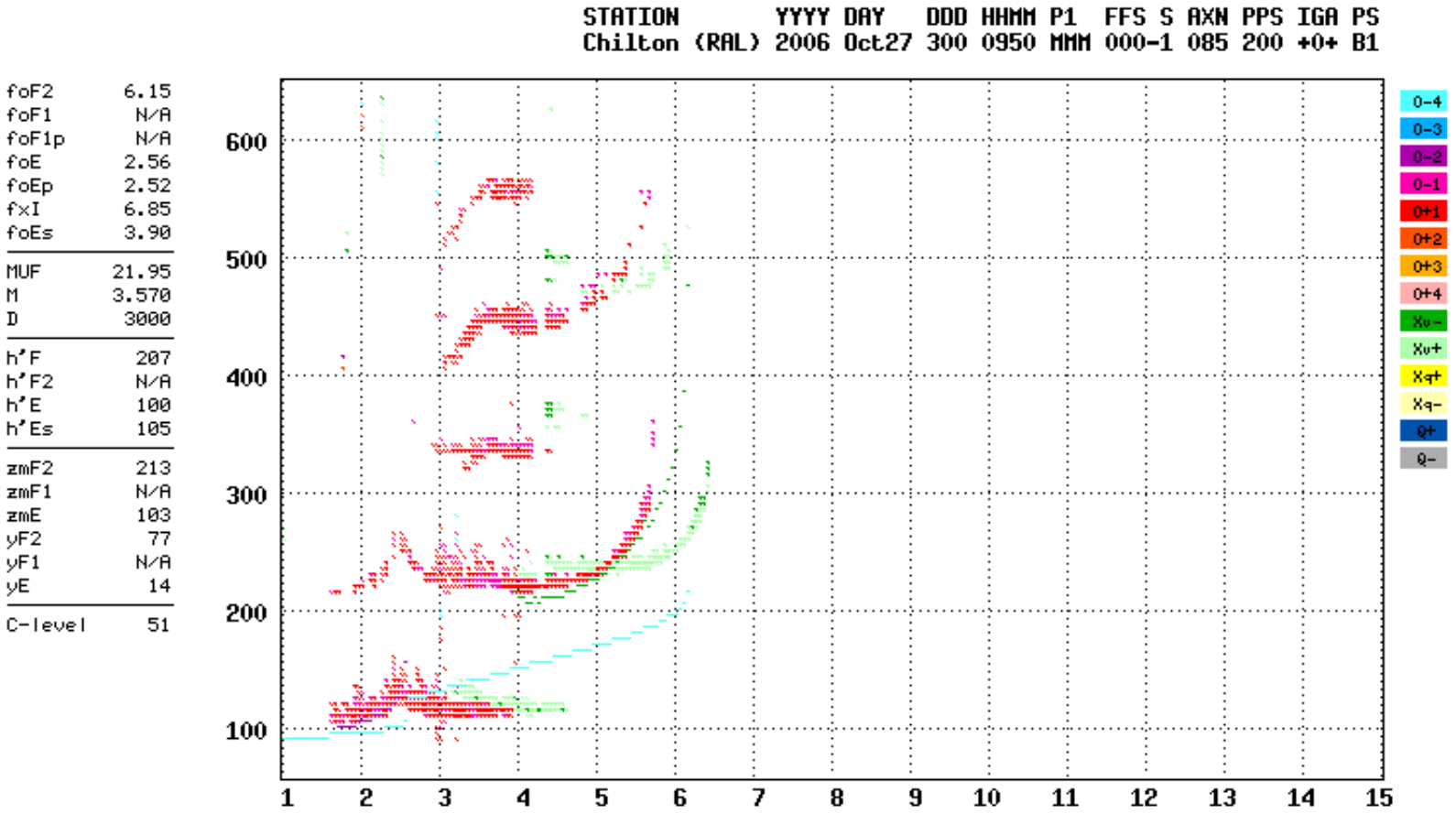
As part of the validation the CASPAR tested simulated the following:

- **hardware changes**
- **software changes**
- **changes in the environment (including legal framework)**
- **changes to the knowledge bases of the Designated Communities**

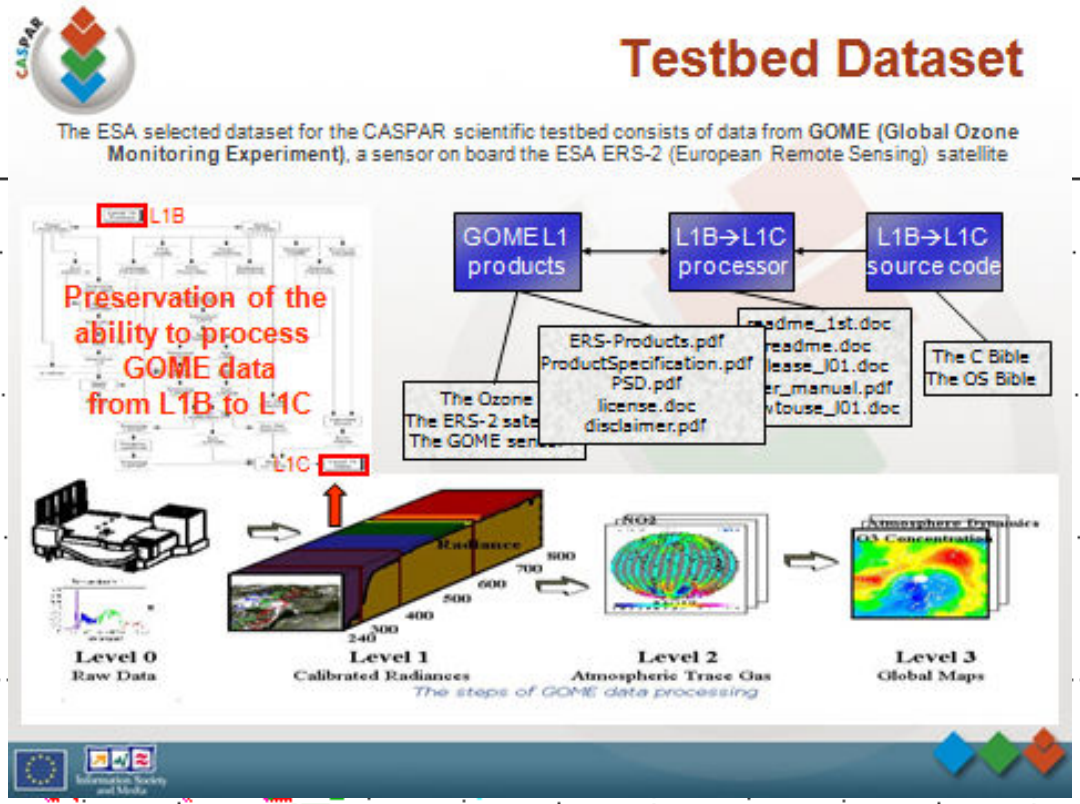
Test scenarios vs Threats to digital preservation

Threat	STFC	ESA	UNESCO	IRCAM	UnivLeeds	CIANT	INA
Users may be unable to understand or use the data e.g. the semantics, format, processes or algorithms involved	☑		☑	☑	☑	☑	
Non-maintainability of essential hardware, software or support environment may make the information inaccessible	☑	☑	☑	☑		☑	
The chain of evidence may be lost and there may be lack of certainty of provenance or authenticity	☑			☑			☑
Access and use restrictions may make it difficult to reuse data, or alternatively may not be respected in future							☑
The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future	☑						

STFC Testbed – various STP



ESA testbed



foF2	6.15
foF1	N/A
foF1p	N/A
foE	2.56
foEp	2.52
fxI	6.85
foEs	3.90
MUF	21.95
M	3.570
D	3000
h'F	207
h'F2	N/A
h'E	100
h'Es	105
zmF2	213
zmF1	N/A
zmE	103
yF2	77
yF1	N/A
yE	14
C-level	51

AXN PPS IGA PS
085 200 +0+ B1

0-4
0-3
0-2
0-1
0+1
0+2
0+3
0+4
Xo-
Xo+
Xq+
Xq-
Q+
Q-

600
500
400
300
200
100

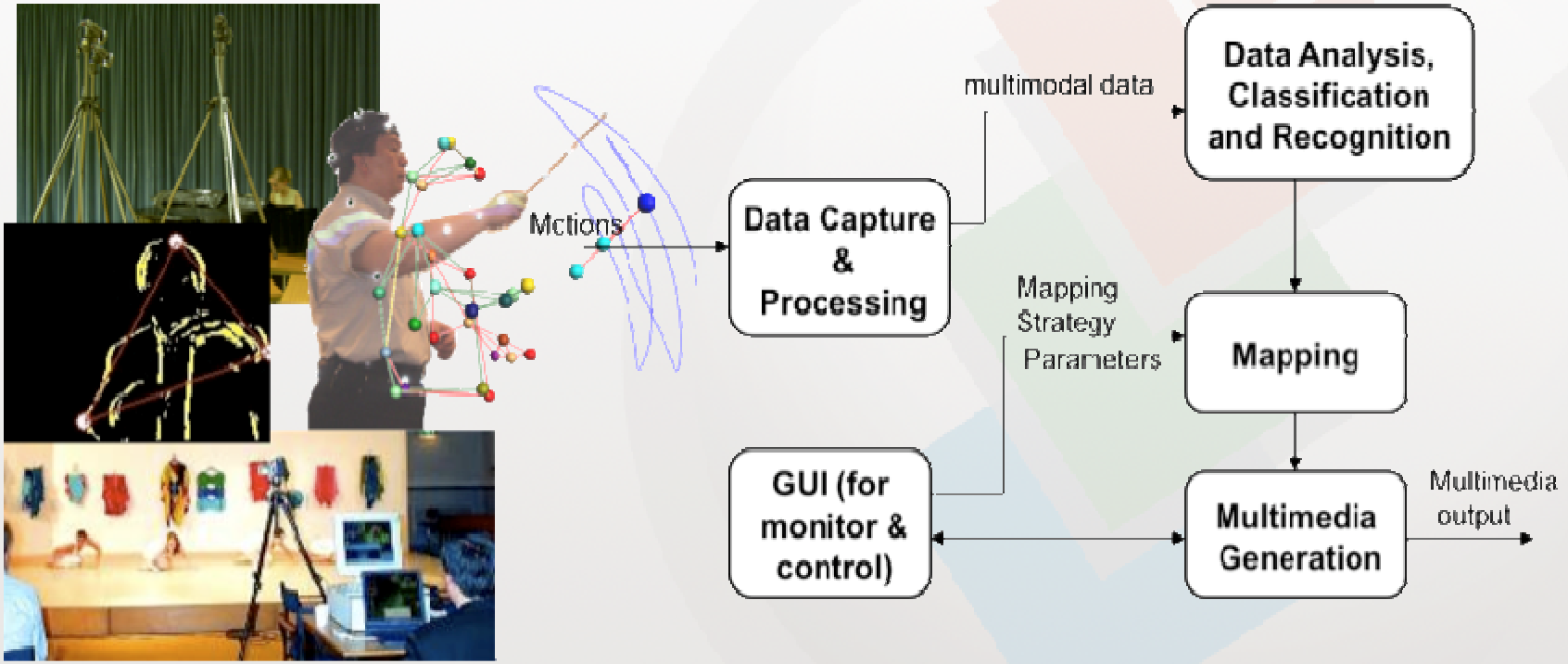
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

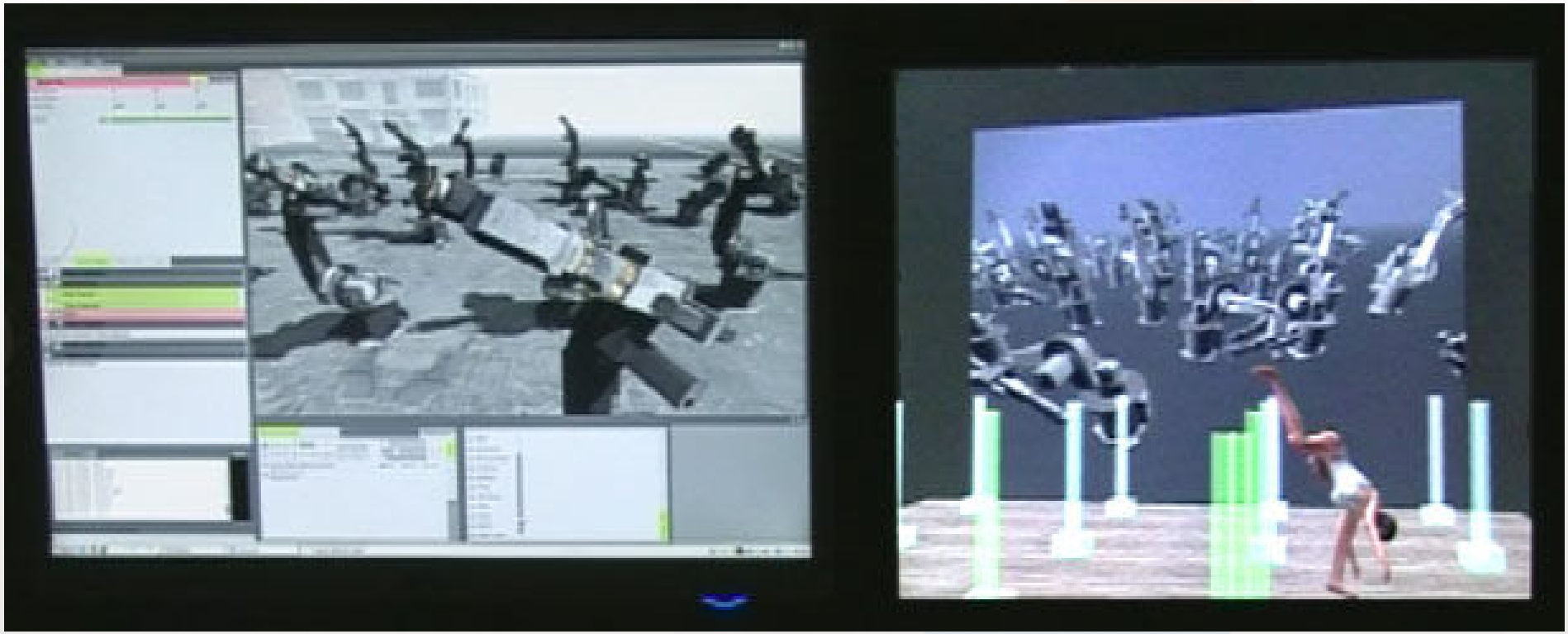
UNESCO testbed

The Villa Livia dataset is a collection of files used within the "virtual museum of the ancient Via Flaminia" project: a 3D reconstruction of several archaeological sites along the ancient Via Flaminia, the largest of them being Villa Livia

This is an elevation grid (height map) of the area where Villa Liva is located. It is an ASCII file in the ESRI GRID file format

Contemporary Art Testbed





Performance Viewer: side-by-side comparison and validation of the transformation. From left to right: 3D visualization in Ogre3D, 3D model of the stage including the virtual dancer in VRML.

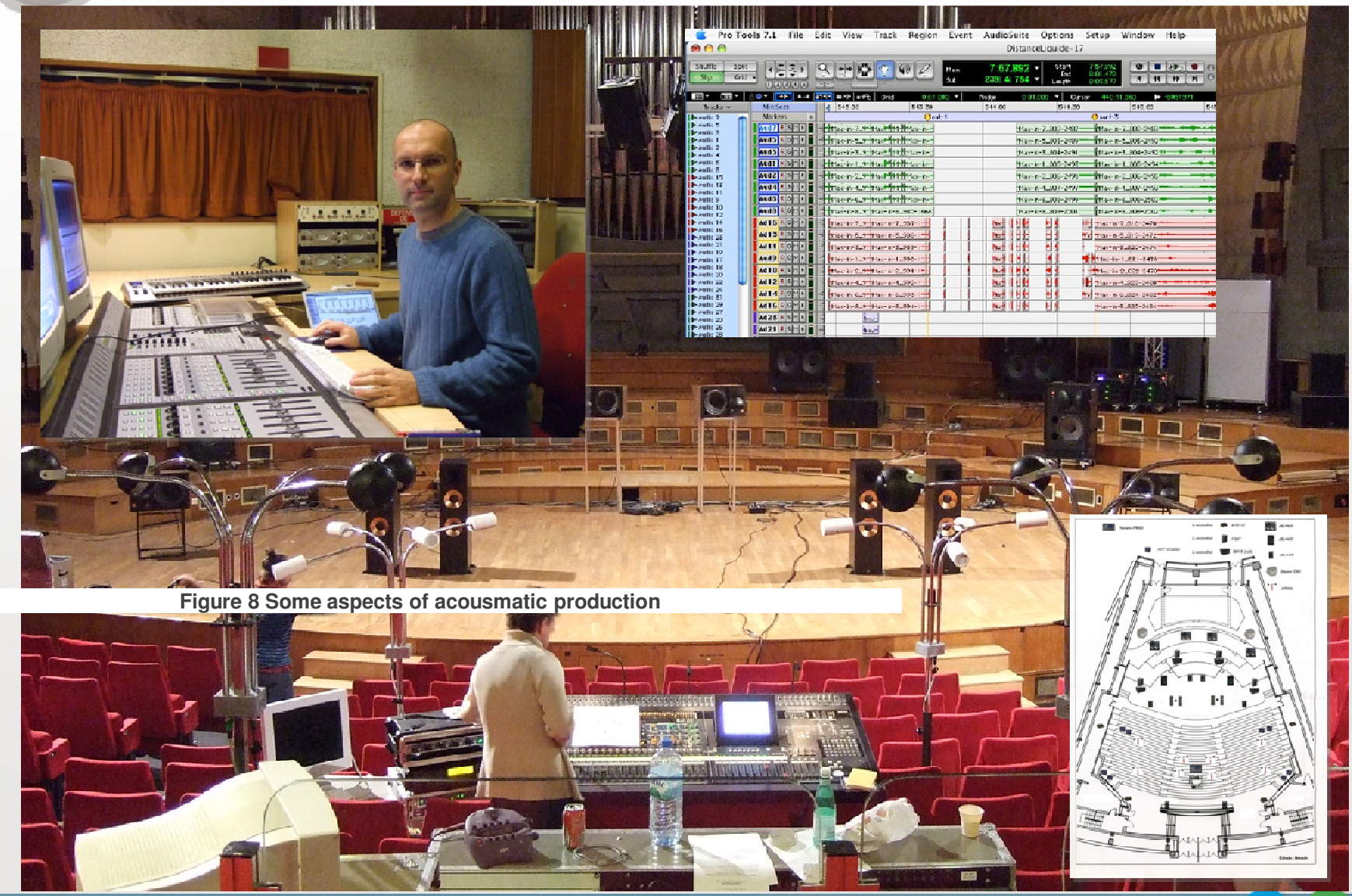


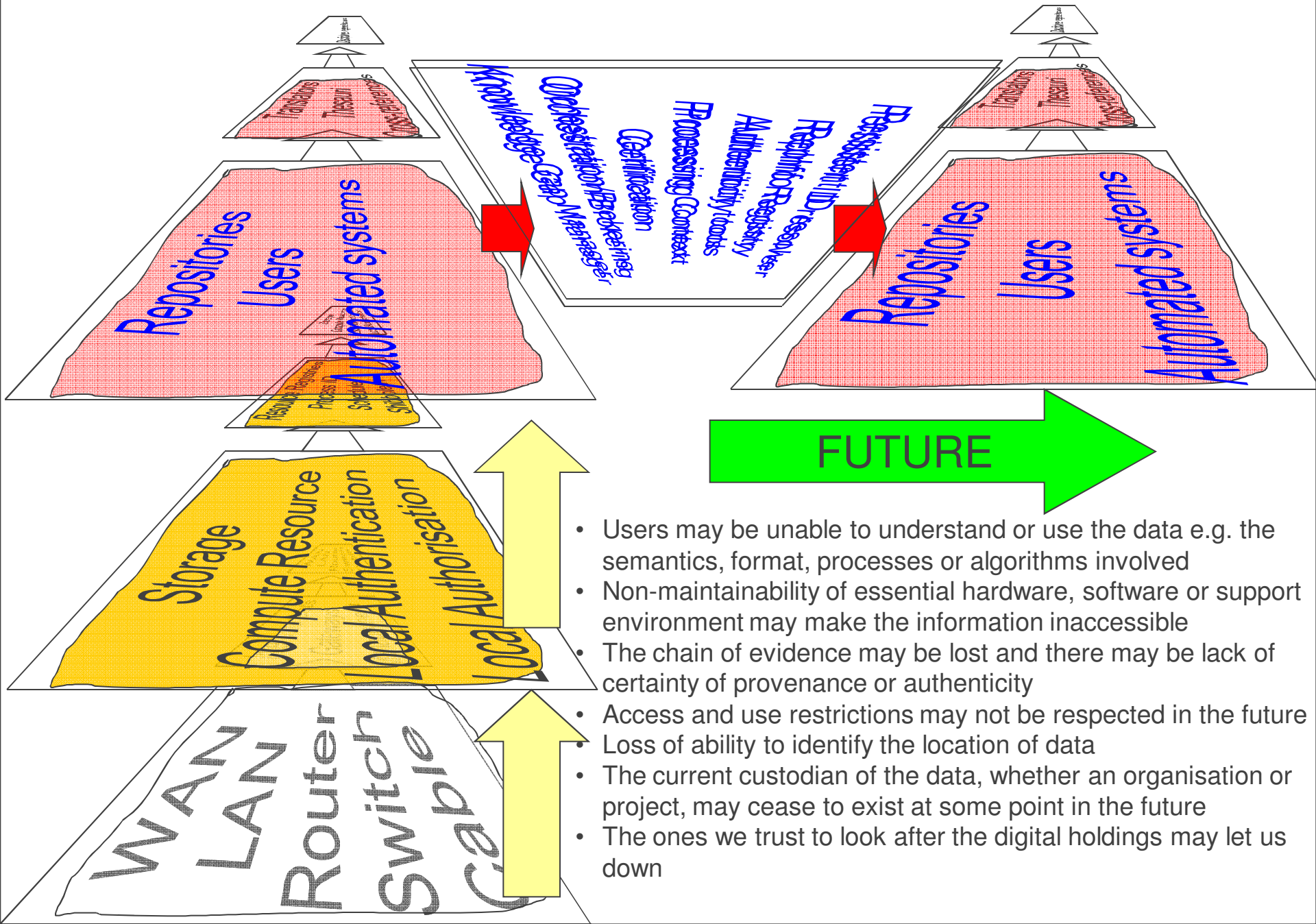
Figure 8 Some aspects of acousmatic production

CASPAR Validation

- In all cases members of the Designated Community, with appropriate changes to mimic changes over time, verified that the metadata was adequate for the use despite simulated changes of hardware, software, environment and Designated Community over time.
- Full details are available in the validation report (CASPAR Validation report, 2009)

Links

- **CASPAR** – http://www.casparpreserves.eu
- **CASPAR Source code** - http://sourceforge.net/projects/digitalpreserve/
- **OAIS Reference Model** - http://public.ccsds.org/publications/archive/650x0b1.pdf
- and the updated draft is available from http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Overview.aspx
- **CASPAR Validation report** http://www.casparpreserves.eu/Members/cclrc/Deliverables/caspar-validation-evaluation-report/at_download/file
- PARSE.Insight:
  - ***www.parse-insight.eu***
- Alliance for Permanent Access:
  - ***www.alliancepermanentaccess.eu***
- Digital Curation Centre:
  - ***www.dcc.ac.uk***



- Users may be unable to understand or use the data e.g. the semantics, format, processes or algorithms involved
- Non-maintainability of essential hardware, software or support environment may make the information inaccessible
- The chain of evidence may be lost and there may be lack of certainty of provenance or authenticity
- Access and use restrictions may not be respected in the future
- Loss of ability to identify the location of data
- The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future
- The ones we trust to look after the digital holdings may let us down

END

Alliance for Permanent Access

CCSDS.ORG
The Consultative Committee for Space Data Systems

DCC

PARSE insight