



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

Stewardship and Preservation Perspectives

ESIP meeting, 2010, Knoxville
David Giaretta

← Alliance for Permanent Access →



Information Society
and Media





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

Overview

- **Audit and Certification overview**
- Metrics
- Digital preservation
- Evidence of preservation
- Audit and Certification Organisation



Information Society
and Media





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

Digital Preservation...

- Easy to do...
- ...as long as you can provide money forever
- Easy to test claims about repositories...
- ...as long as you live a long time



Information Society
and Media

DigCCurr 2009:
Digital Curation





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

MONEY



Information Society
and Media





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

History of Audit & Certification of Digital Repositories

- Demand for some method of evaluating digital repositories from mid-90's
- Section of OAIS on roadmap for development of related standards includes an item for accreditation of archives
- TRAC (Trustworthy Repositories Audit & Certification: Criteria and Checklist)
 - RLG and NARA led this work as agreed by OAIS group
 - Based on OAIS Reference Model and Report on Trusted Digital Repositories: Attributes and Responsibilities (RLG-OCLC, 2002)
 - Not really up to ISO audit & certification



Information Society
and Media

7/26/10

5



CCSDS Working Group

- Charter of WG agreed at meeting Jan 2007
- Goal: Obtain ISO approval of a standard that establishes the criteria that a repository/archive must meet to be designated an ISO Trusted Digital Repository
 - on which a full audit and certification of digital repositories can be based
- Following route of OAIS
 - CCSDS is the “working arm” of TC20/SC13 of ISO
- Based on TRAC document



Current status

- Two standards documents prepared
 - Metrics for Audit and Certification of Digital Repositories
 - Requirements for Bodies Providing Audit and Certification of Digital Repositories
- Both now submitted to CCSDS/ISO
- Completely public process



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

Overview

- Audit and Certification overview
- **Metrics**
- Digital preservation
- Evidence of preservation
- Audit and Certification Organisation



Information Society
and Media





The metrics document

Section A: Organisational Infrastructure

Section B: Digital Object Management

Section C: Infrastructure and Security Risk
Management

Metrics and their structure:

- Statement of requirement
- Supporting text
- Examples of Ways the Repository can Demonstrate it is Meeting this Requirement
- Discussion



The metrics document

3.2.1.2 The repository shall have the appropriate number of staff to support all functions and services.

Supporting Text

This is necessary in order to ensure repository staffing levels are adequate for preserving the digital content and providing a secure, quality repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Organizational charts; definitions of roles and responsibilities; comparison of staffing levels to industry benchmarks and standards.

Discussion

The repository should determine the appropriate number and level of staff that corresponds to requirements and commitments. The repository should also demonstrate how it evaluates staff effectiveness and suitability to support its functions and services.



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

Practical audit and certification

- Depends upon the experience of the auditors
 - Metrics provide guidance on questions to ask
 - Primary TDR Accreditation Board is first set of auditors
- ISO Audit & Certification process involves cycles of improvement
- Certification not “YES/NO”
 - More like “OK BUT CAN IMPROVE HERE”
- Broad, common, understanding of digital preservation is needed by auditors.



Information Society
and Media





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

Overview

- Audit and Certification overview
- Metrics
- **Digital preservation**
- Evidence of preservation
- Audit and Certification Organisation



Information Society
and Media





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

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 preservation
 - Defines terminology and compliance criteria



Information Society
and Media



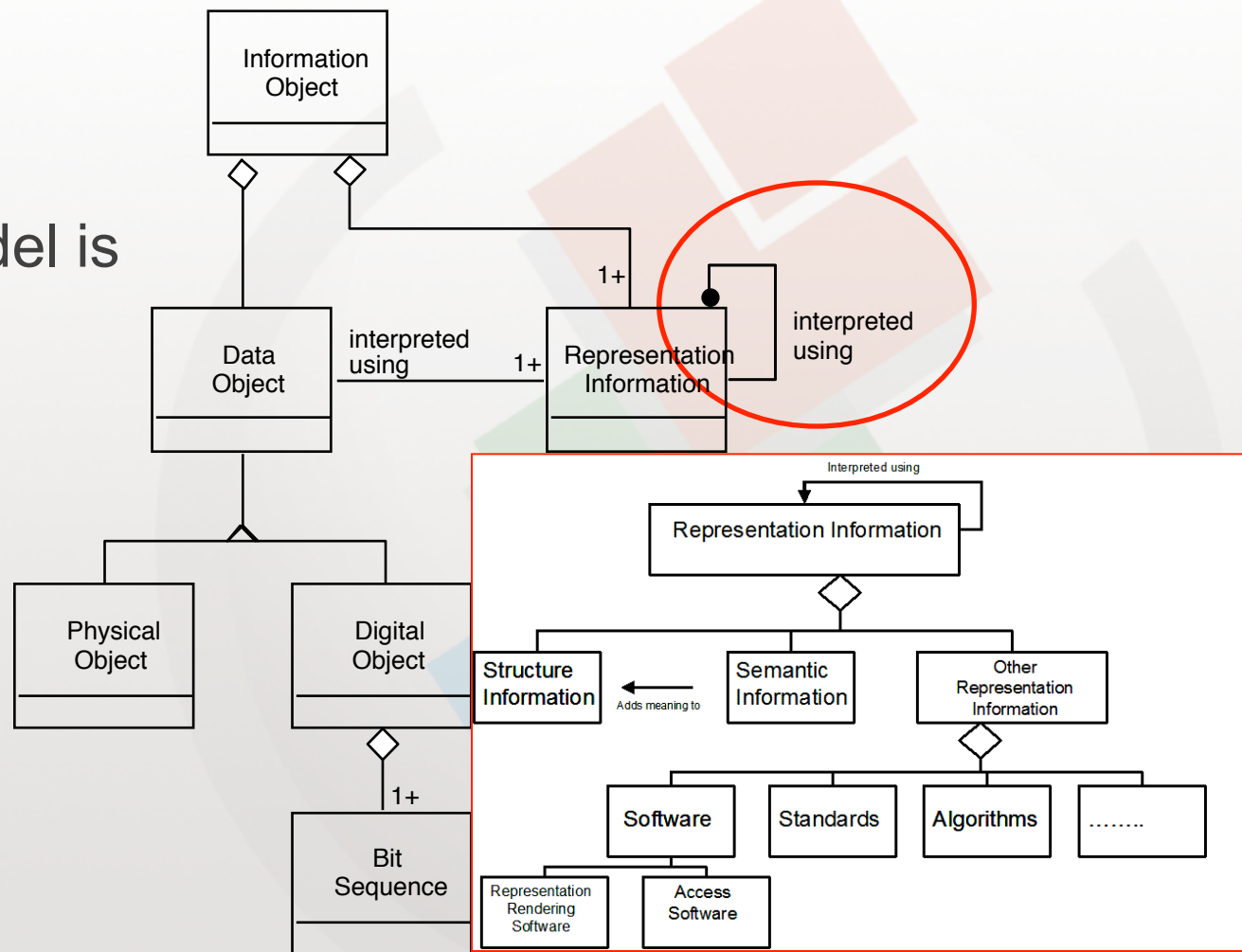


Information Model & Representation Information

The Information Model is
key

Recursion ends at
KNOWLEDGEBASE of the
DESIGNATED COMMUNITY

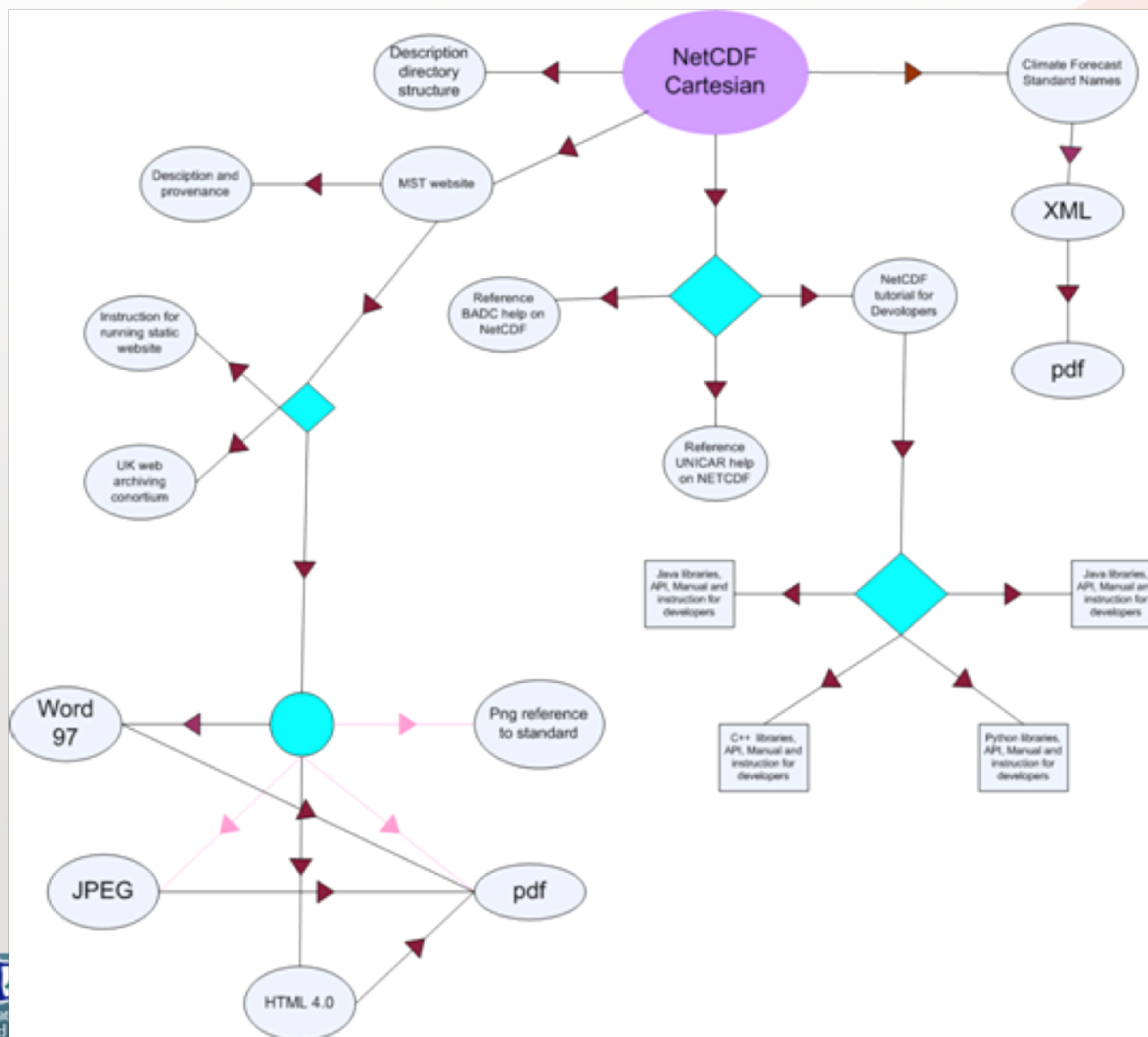
(this knowledge will change
over time and region)





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

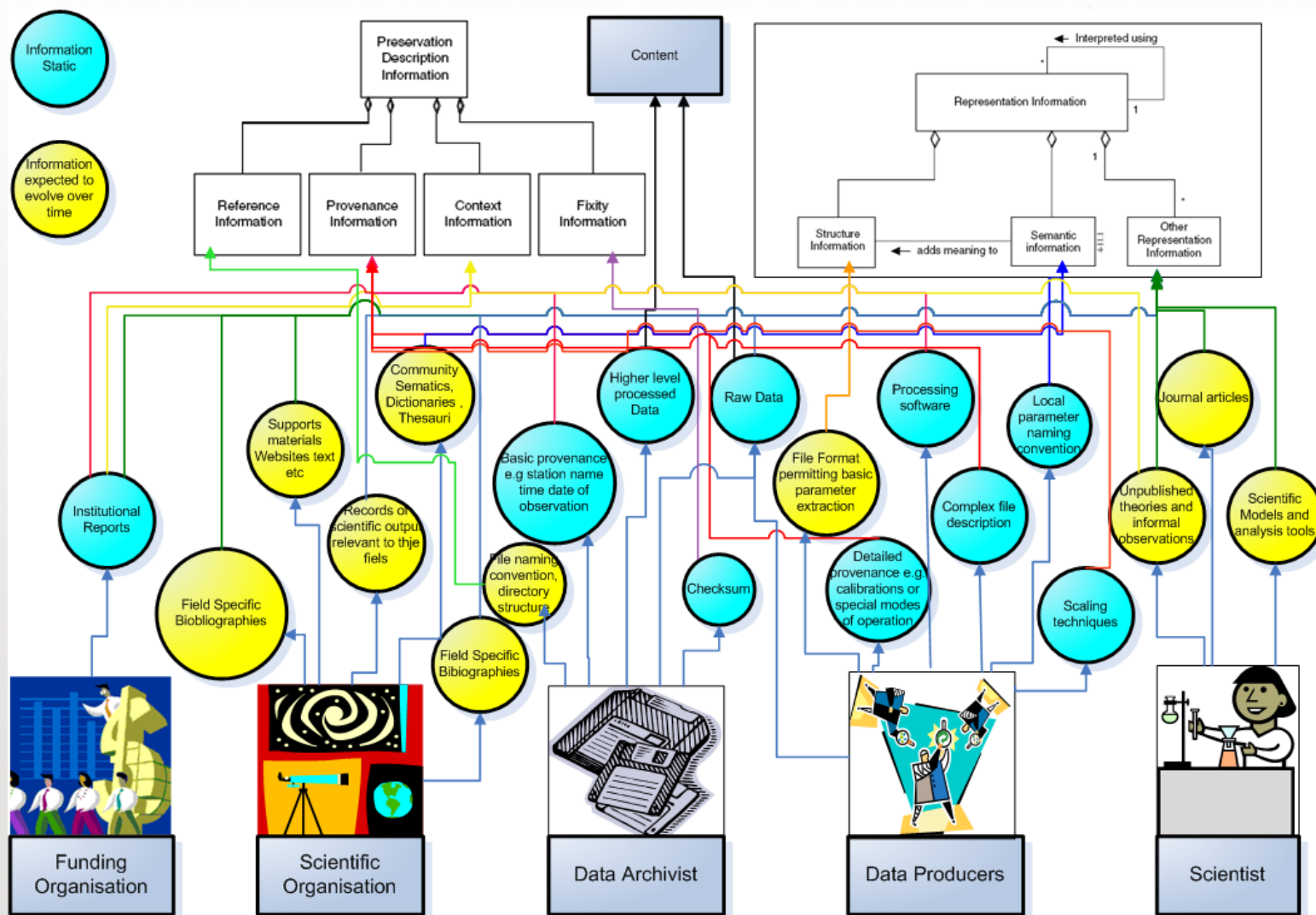
Representation Information Network





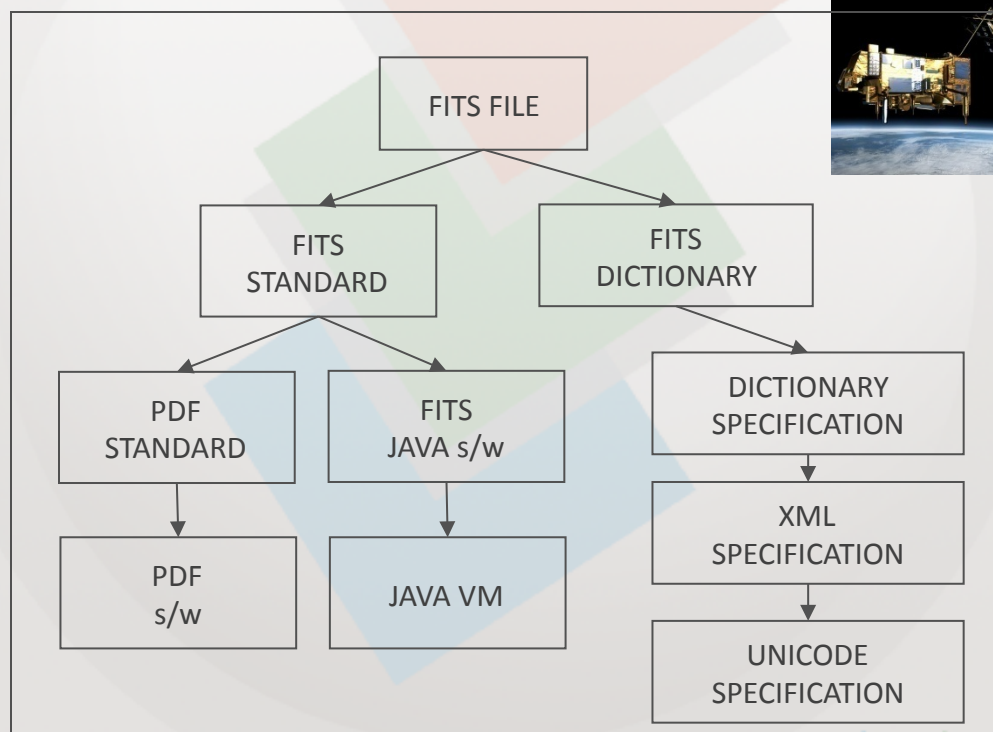
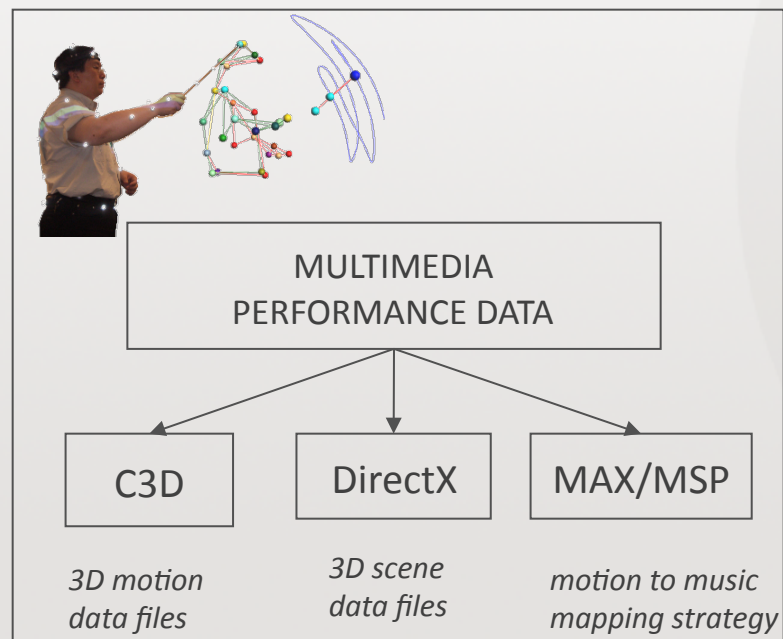
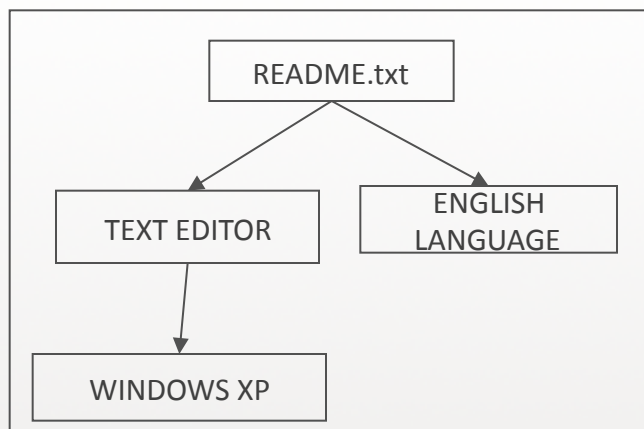
Cultural, Artistic and Scientific knowledge
for Preservation, Access and Retrieval

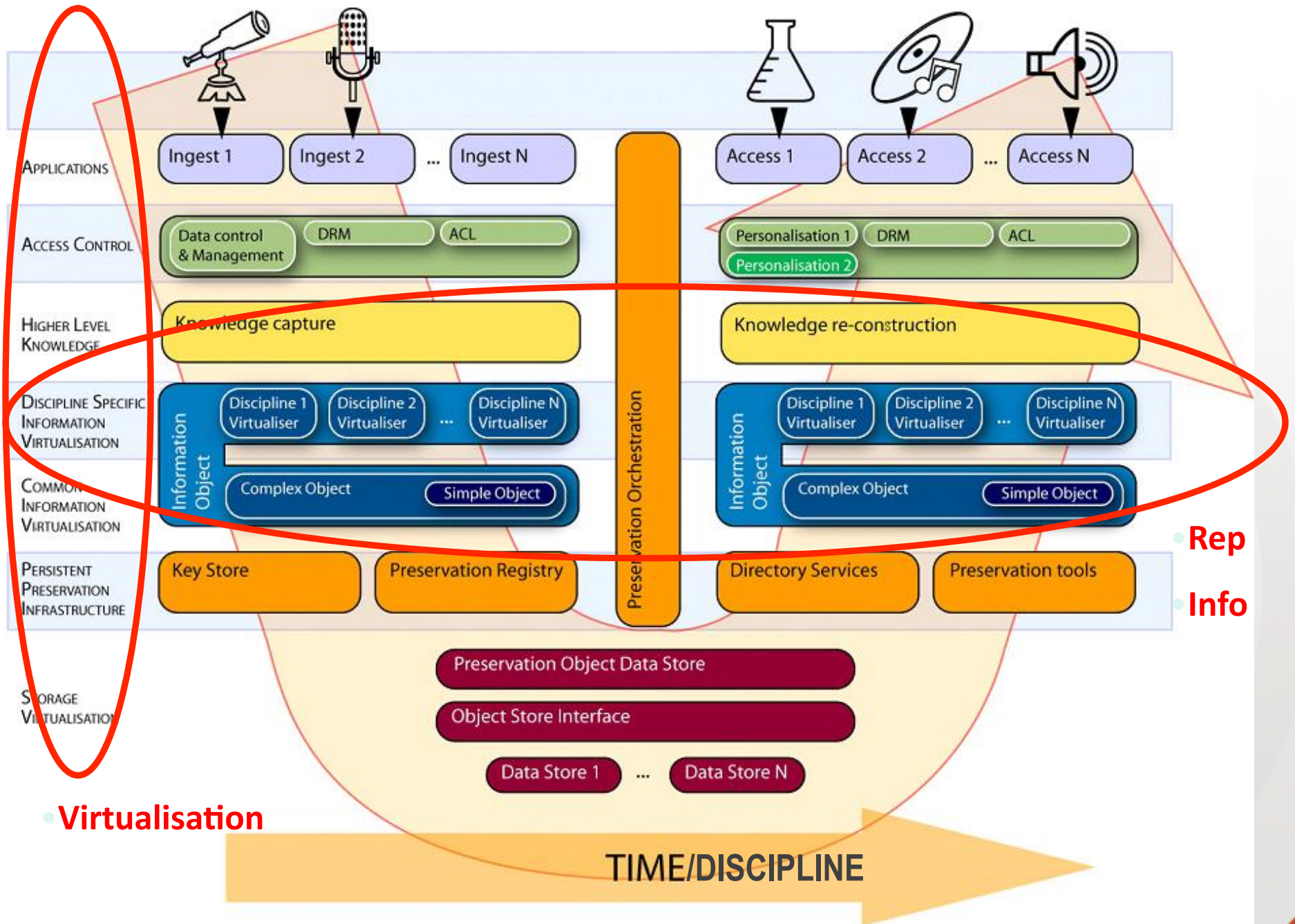
Preservation Data Flows and Strategies





Modules and Dependencies: defining the Designated Community







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

Basic concept of CASPAR

- Digital preservation had been dominated by libraries and (state) archives
 - However there was a focus there on “rendered objects” and “**metadata**”
 - Tendency to think data is an “easy” add-on
- HOWEVER
- Need to deal with DATA – processed to new things, not just rendered
 - Need to follow OAIS – finer grained view
 - Need to test and prove that things work



Information Society
and Media





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

General Surveys to stakeholders

Funding/policy
< responses

Research
1397 responses

Plus a similar number from in-
depth case studies

Data management
(preservation)
273 responses

Publishing
186 responses

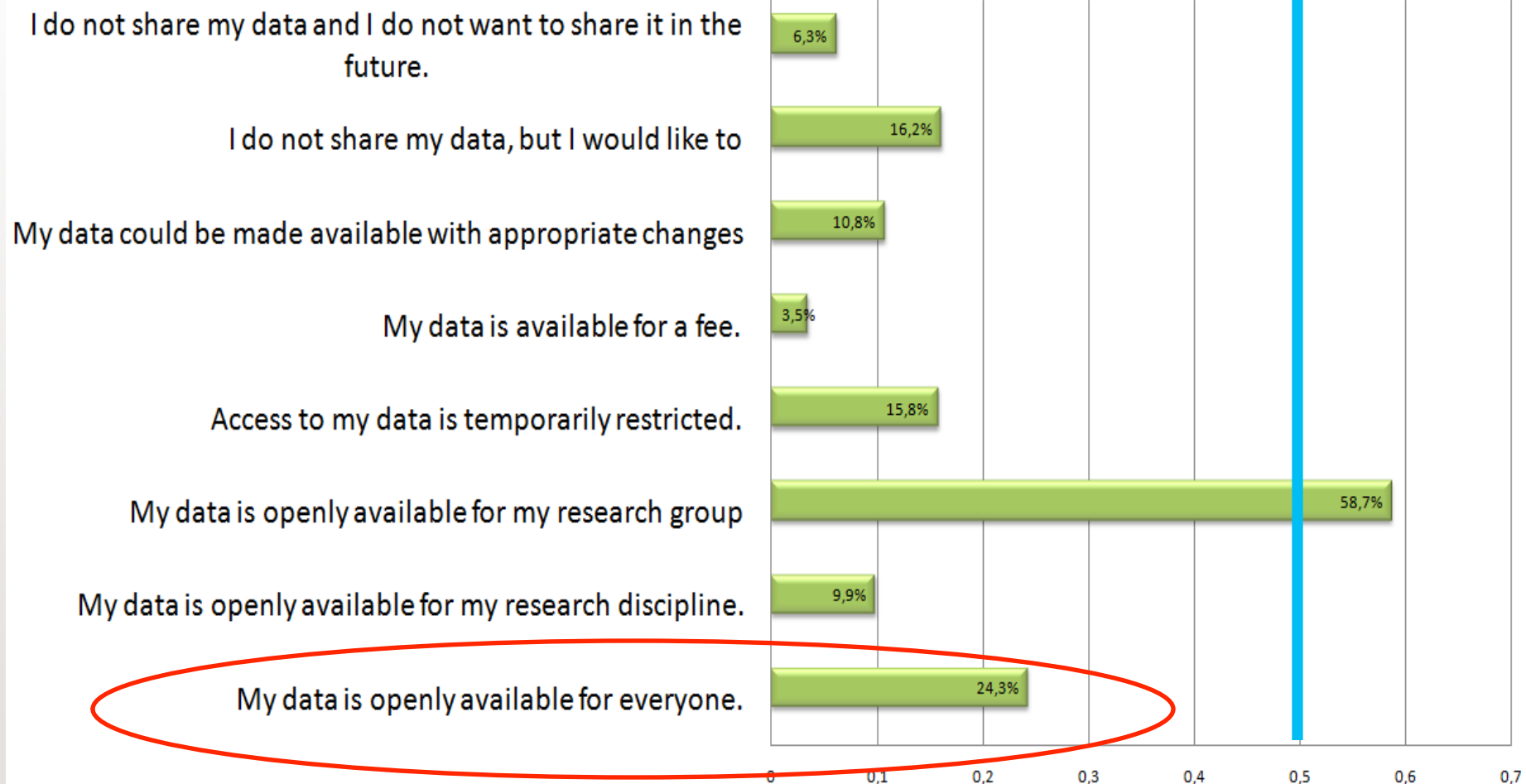


Information Society
and Media





Sharing of data (R)

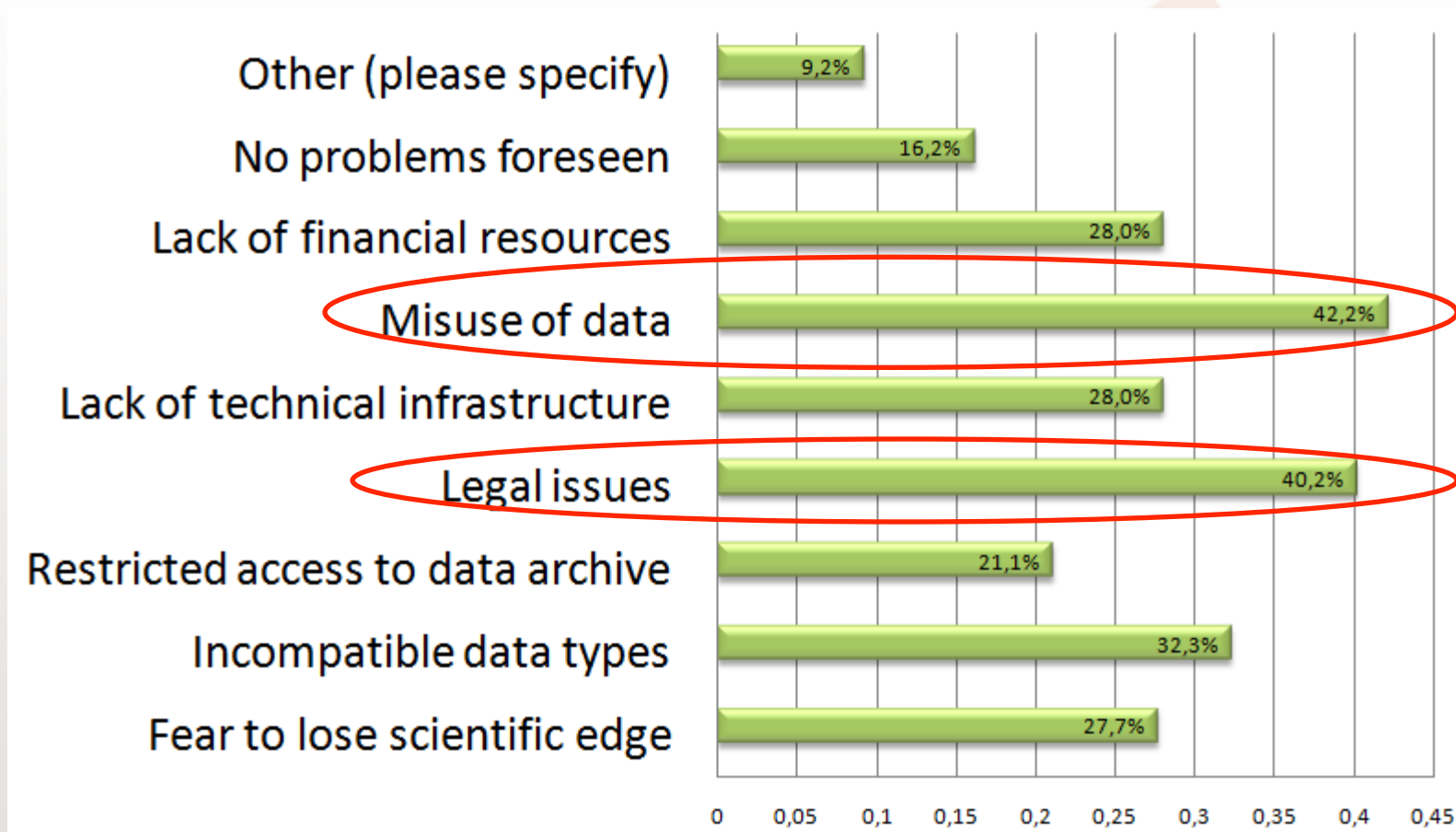


How open is your data?





Sharing of data (R)



Which constraints do you see in making data open?





Threats to preservation

1. The ones we trust to look after the digital holdings may let us down.
2. The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future.
3. Loss of ability to identify the location of data.
4. Access and use restrictions (e.g. Digital Rights Management) may not be respected in the future.
5. Evidence may be lost because the origin and authenticity of the data may be uncertain.
6. Lack of sustainable hardware, software or support of computer environment may make the information inaccessible.
7. Users may be unable to understand or use the data e.g. the semantics, format or algorithms involved.





Threats to preservation (R)

The ones we trust to look after the digital holdings may let us down

The current custodian of the data may cease to exist

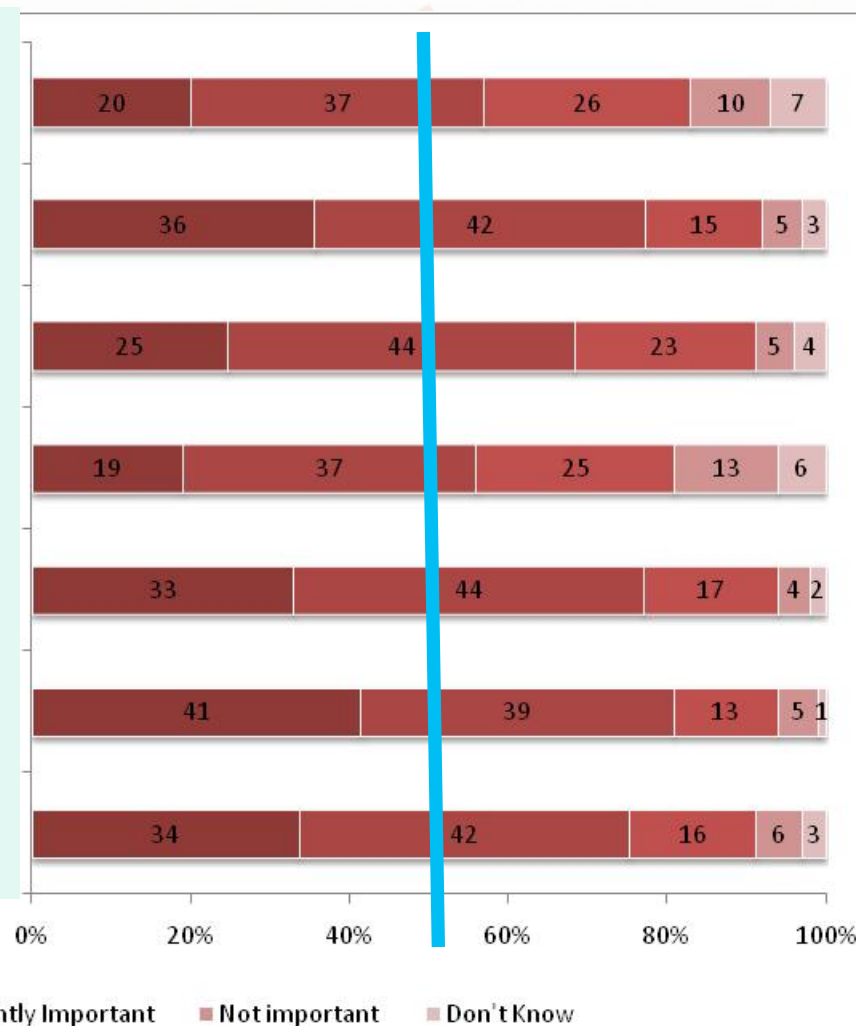
Loss of ability to identify the location of data

Access and use restrictions may not be respected in the future

Evidence may be lost

Lack of sustainable hardware/software

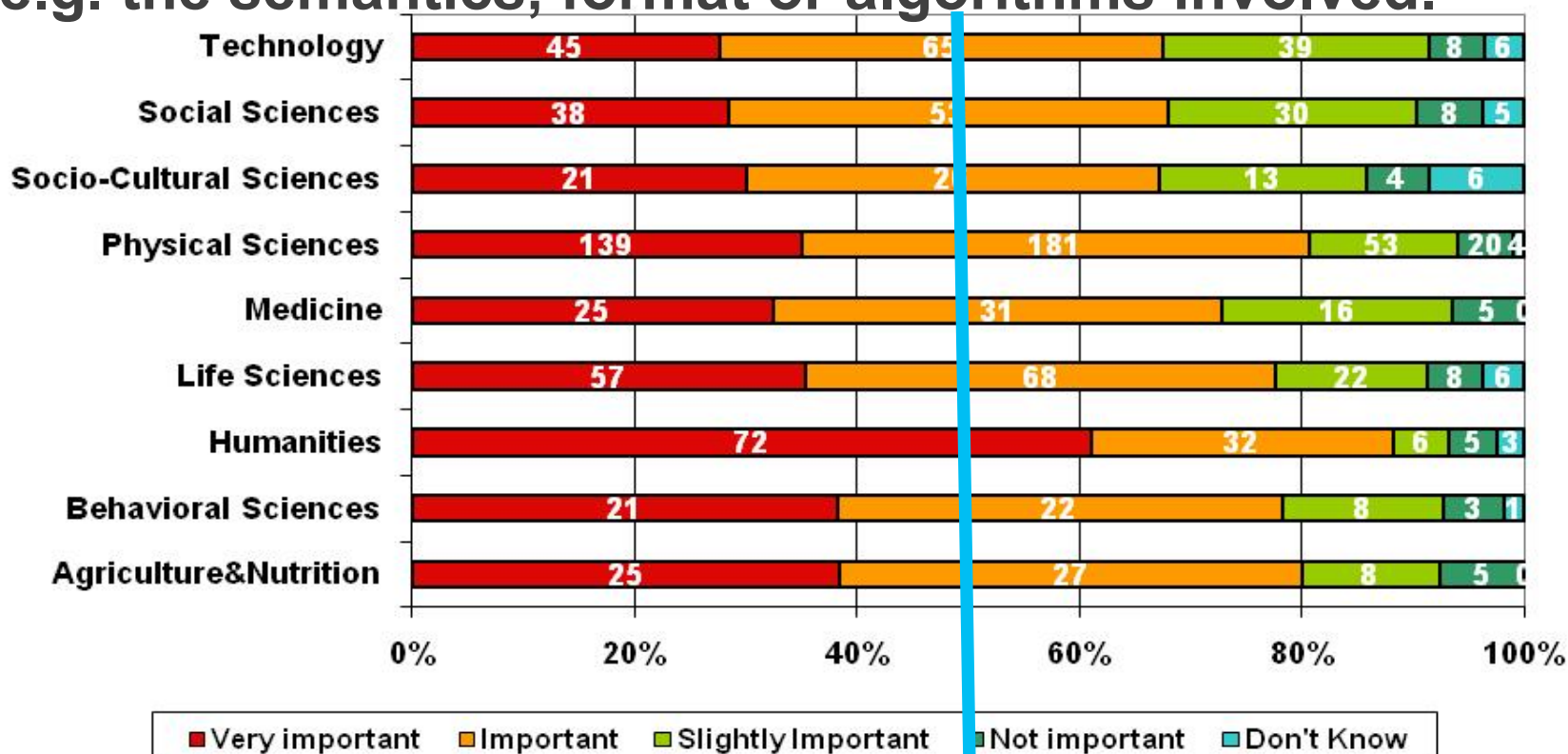
Users may be unable to understand or use the data





Threats to preservation (R)

Users may be unable to understand or use the data
e.g. the semantics, format or algorithms involved.



Threat	Examples
Users may be unable to understand or use the data e.g. the semantics, format, processes or algorithms involved	Things which used to be tacit knowledge are no longer known. For example particular terminology may fall out of use, whole languages may die, paradigms of ways to analyse problems may disappear.
Non-maintainability of essential hardware, software or support environment may make the information inaccessible	Hardware on which one currently depends, for example Intel x86 CPUs, or tape readers, or whole operating systems which software relies on may no longer function through lack of support. Open Source software may be available but its developers may drift away.
The chain of evidence may be lost and there may be lack of certainty of provenance or authenticity	Someone may claim that a digital object is something of significance, for example a diary of a famous person but one may have doubts about its origin and whether it has been surreptitiously altered.
Access and use restrictions may make it difficult to reuse data, or alternatively may not be respected in future	Your own data, which you had submitted to a repository, may be used without your permission even though you explicitly stated that it should be kept for 30 years without anyone accessing it.
Loss of ability to identify the location of data	A Web page contains a link to an image but the URL does not work – in fact the DNS may say there is no such machine registered.
The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future	The organisation that is charged with looking after the digital object may lose its funding.
The ones we trust to look after the digital holdings may let us down	The people we entrust with our digital objects may make preservation decisions which in the long run mean that the digital objects are not usable.



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

Overview

- Audit and Certification overview
- Metrics
- Digital preservation
- **Evidence of preservation**
- Audit and Certification Organisation



Information Society
and Media





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

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



Information Society
and Media





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	J		J	J	J	J	
Non-maintainability of essential hardware, software or support environment may make the information inaccessible	J	J	J	J		J	
The chain of evidence may be lost and there may be lack of certainty of provenance or authenticity	J			J			J
Access and use restrictions may make it difficult to reuse data, or alternatively may not be respected in future							J
The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future	J						





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

Overview

- Audit and Certification overview
- Metrics
- Digital preservation
- Evidence of preservation
- **Audit and Certification
Organisation**



Information Society
and Media





“Requirements for bodies ...”

- Relation to analagous standards
 - ISO 17021 “Conformity assessment — Requirements for bodies providing audit and certification of management systems”
 - Elaborates on this standard in the context of Trustworthy Digital Repositories
- Structure
 - Follows ISO 17021



“Requirements for bodies ...”

7.2 PERSONNEL INVOLVED IN THE CERTIFICATION ACTIVITIES

The requirements from ISO/IEC 17021:2006, Clause 7.2 apply. In addition, the following TDR audit and certification specific requirements and guidance apply.

7.2.1 COMPETENCE OF CERTIFICATION BODY PERSONNEL

The following training requirements apply to all members of the audit team, with the exception of d), which can be shared among members of the audit team. The certification body shall have criteria for the training of audit team members that ensures

- a) knowledge of CCSDS 652.0-M-1/ISO 16363[1] and other relevant normative documents;
- b) understanding of digital preservation;
- c) understanding of risk assessment and risk management of digitally encoded information;
- d) technical knowledge of the digital preservation aspects which apply to the activity to be audited;
- e) general knowledge of regulatory requirements relevant to TDRs;
- f) knowledge of management systems;
- g) understanding of the principles of auditing based on ISO 19011;



The wider context

- There are other standards/approaches
 - For example “Data Seal of Approval”
- We can envisage a marketplace for audit/certification services, appropriate to the needs of individual repositories





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

European Framework for Audit & Certification of Digital Repositories

- MoU signed which provides
 - **BASIC CERTIFICATION** is granted to repositories which obtain DSA certification;
 - **EXTENDED CERTIFICATION** is granted to Basic Certification repositories which in addition perform a structured, externally reviewed and publicly available self-audit based on ISO 16363 (or DIN 31644);
 - **FORMAL CERTIFICATION** is granted to repositories which in addition to Basic Certification obtain full external audit and certification based on ISO 16363 (or equivalent DIN 31644)

Easy “buy-in”,
few metrics

More metrics

“Gold standard”



Information Society
and Media





Next steps

- Awaiting feedback from CCSDS/ISO
- Set up of “Primary TDR Accreditation Board”
- Test audits to test methodology and practicality
- Start ISO audits 2011
- Set up 3 layer framework in USA and elsewhere?
- Accreditation of more auditors and other certification bodies
- Commercial self-funding audit & certification system (2013?)



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

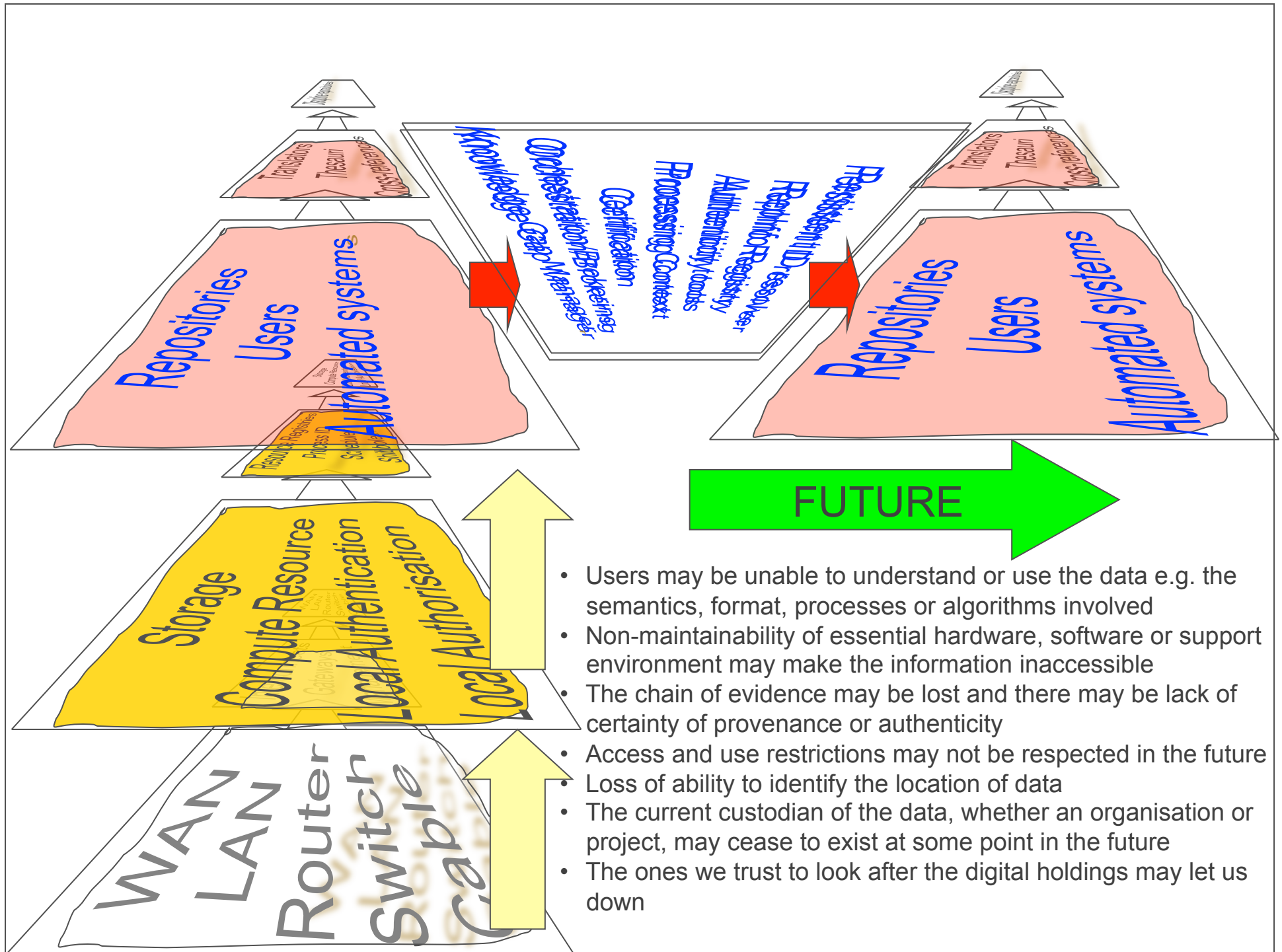
Links

- **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>
- **CCSDS/ISO RAC** - <http://wiki.digitalrepositoryauditandcertification.org>
- **CASPAR** – <http://www.casparpreserves.eu>
- **CASPAR Source code** - <http://sourceforge.net/projects/digitalpreserve>
- **CASPAR Validation report** http://www.casparpreserves.eu/Members/ccirc/Deliverables/caspar-validation-evaluation-report/at_download/file
- **PARSE.Insight:**
 - www.parse-insight.eu
- **Alliance for Permanent Access:**
 - www.alliancepermanentaccess.eu



Information Society
and Media





- 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



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

END

← Alliance for Permanent Access →



The Consultative Committee for Space Data Systems



Information Society
and Media

