

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

David Giaretta

← Alliance for Permanent Access →



The Consultative Committee for Space Data Systems







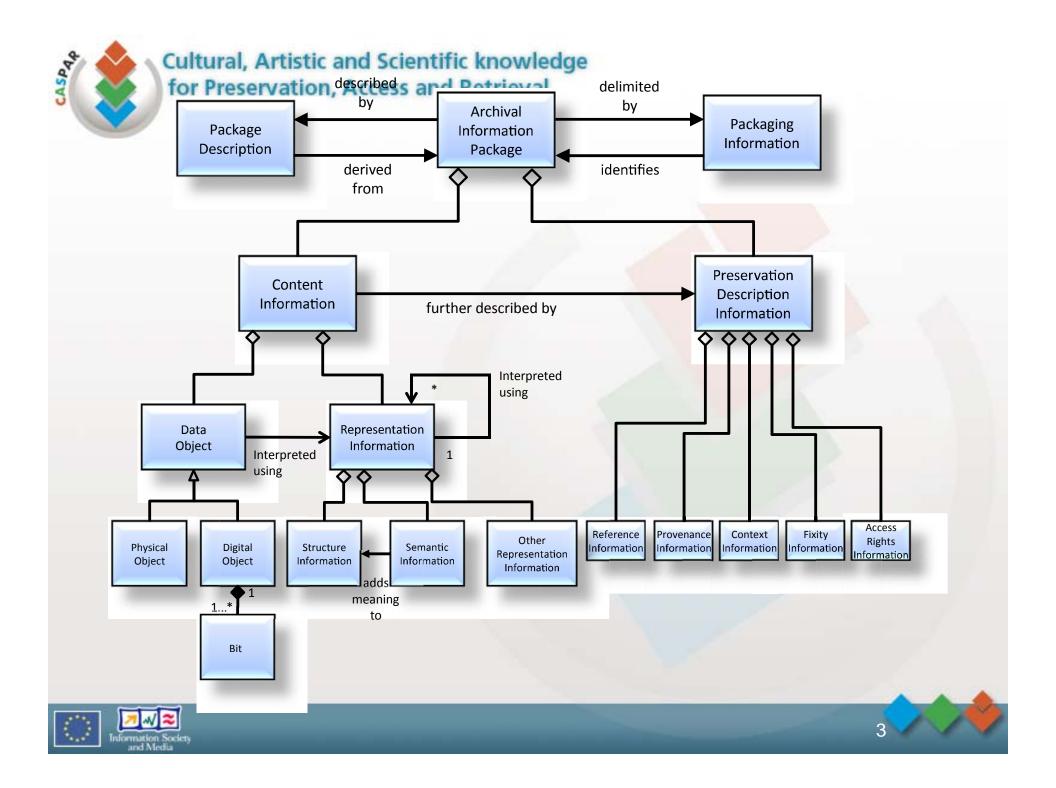


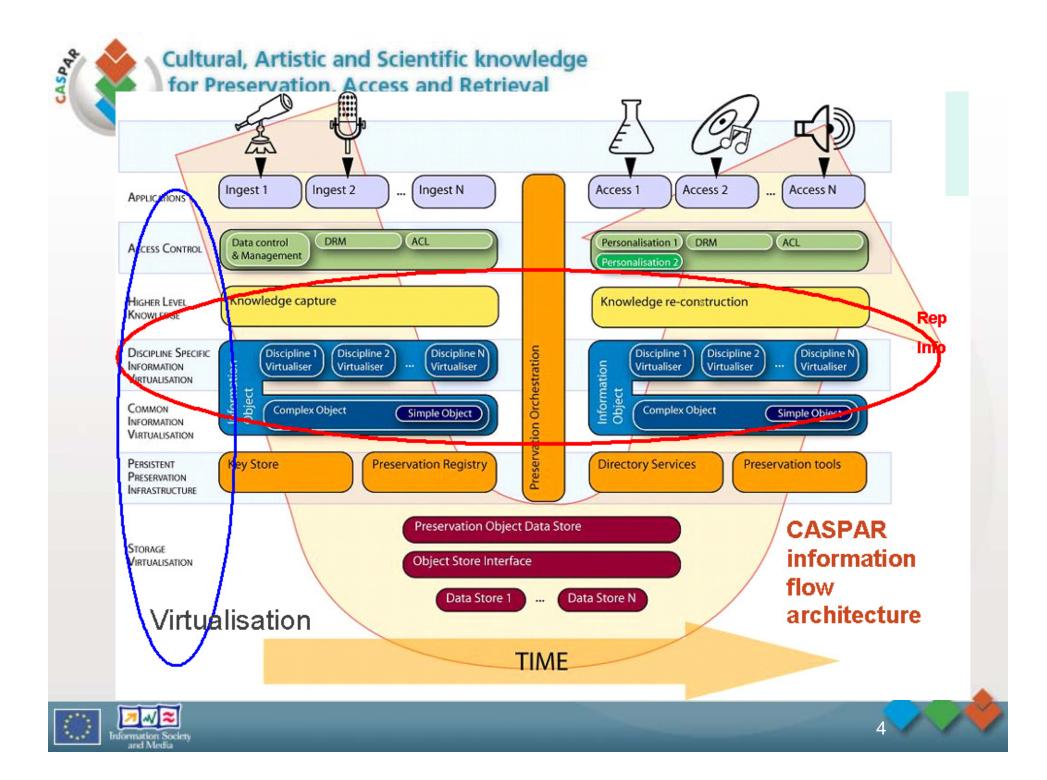


EU FP6 Integrated Project

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











CASPAP

5



RFADMF.txt

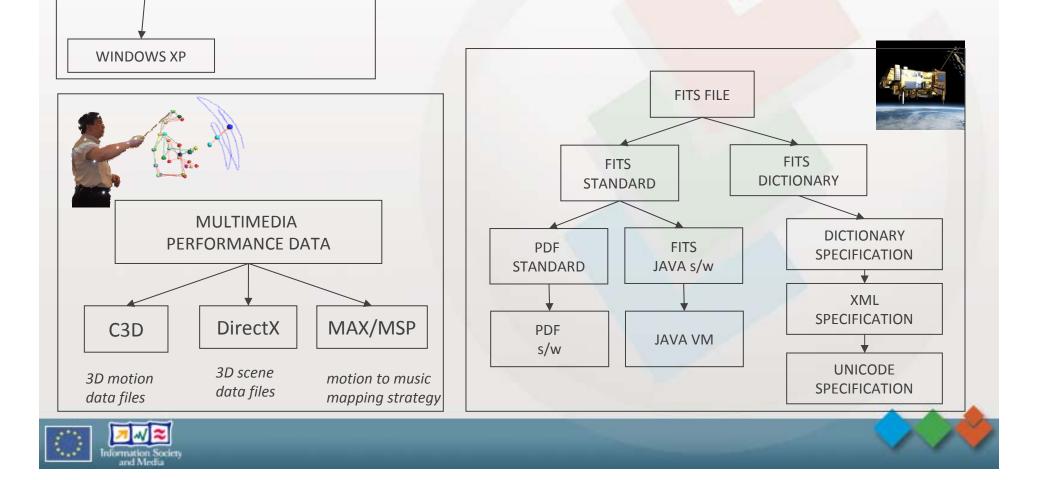
TEXT EDITOR

ENGLISH

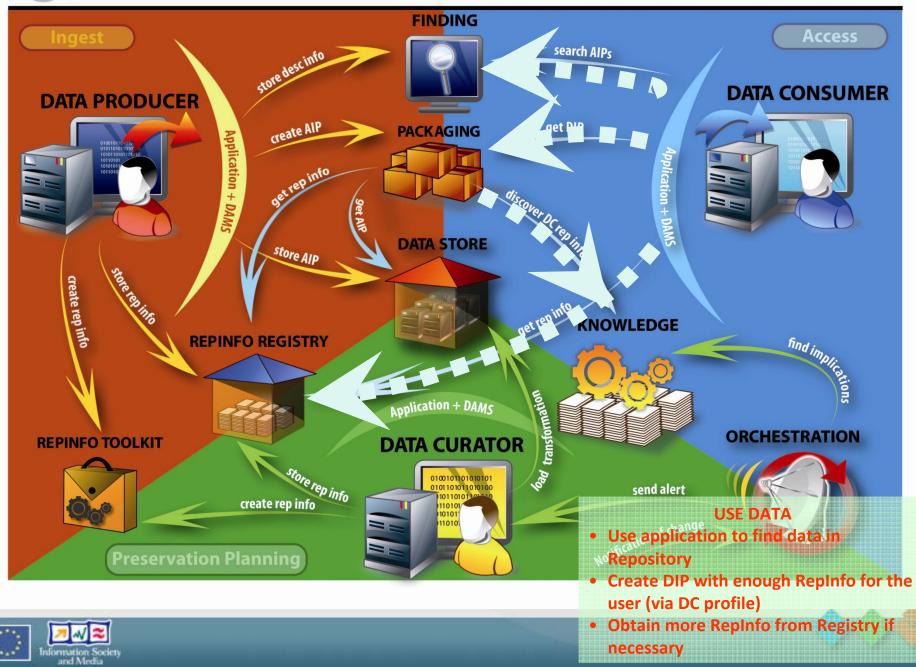
LANGUAGE

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

Modules and Dependencies: defining the Designated Community



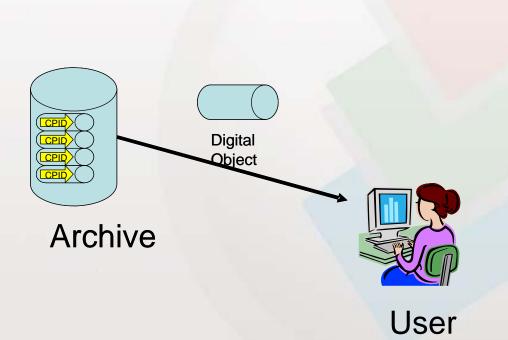






User gets data from archive.

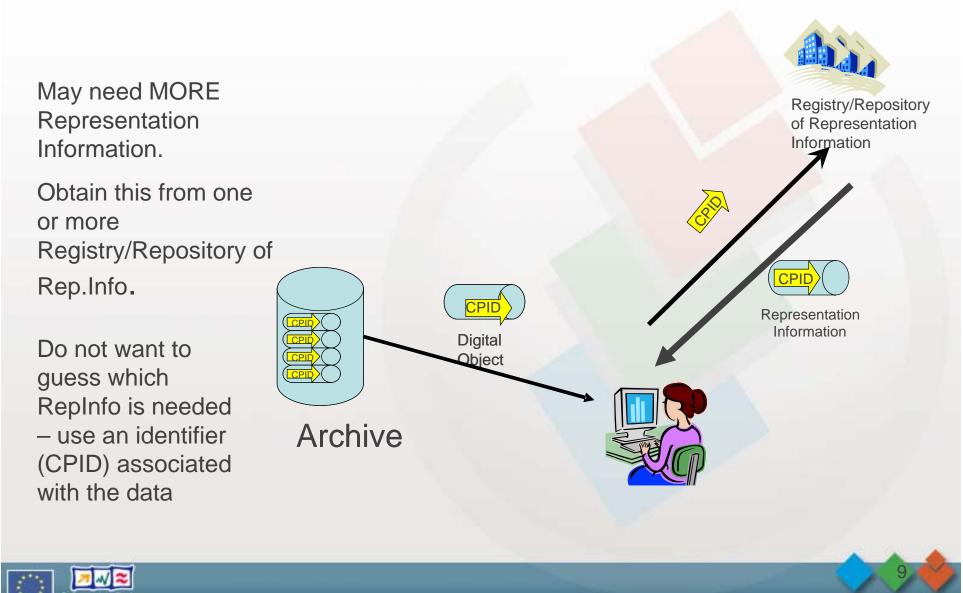
User may be unfamiliar with the data so needs Representation Information. Some may be packaged with the data



The Digital Object could have some RepInfo packed with it.







CASPAR Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval At some point in the future there is a need for further Representation Information. RepInfo Toolkit Someone creates the necessary CPI RepInfo using

whatever tools are

appropriate.



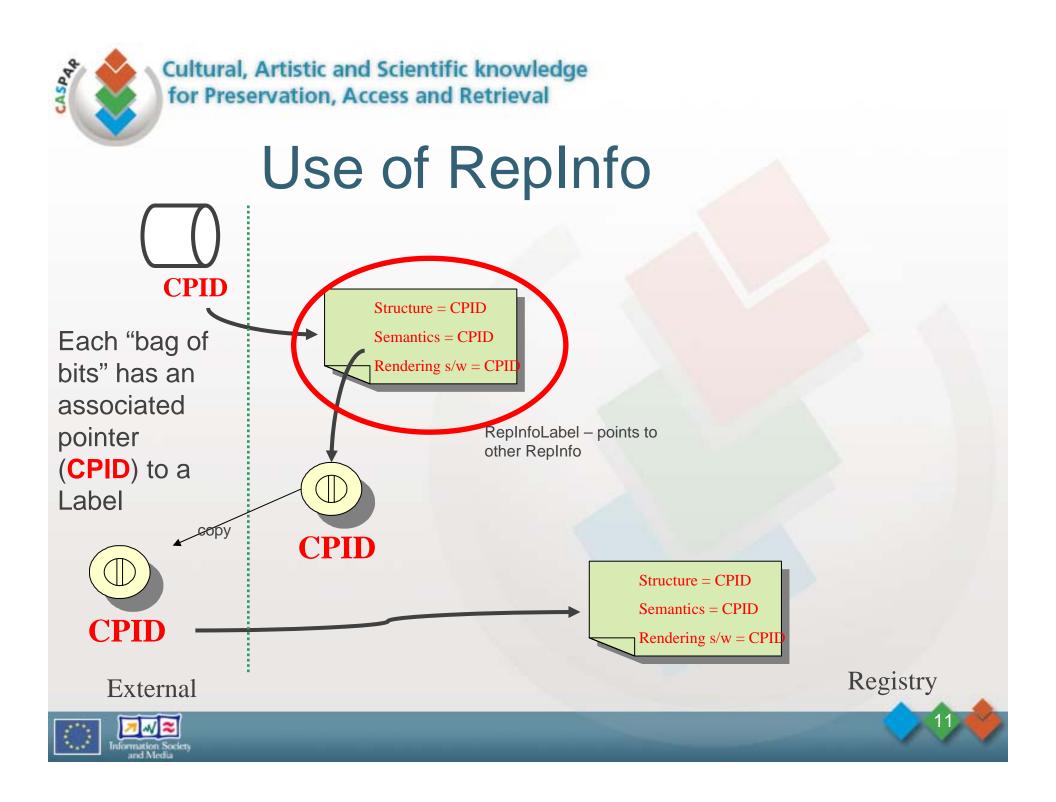
Digital Object Registry/Repository

of Representation

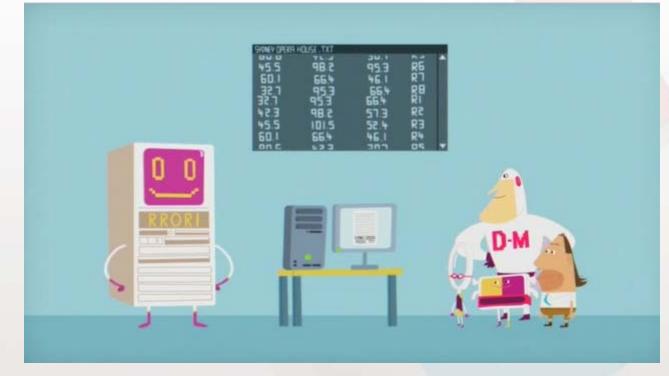
Information

CPID

Representation Information



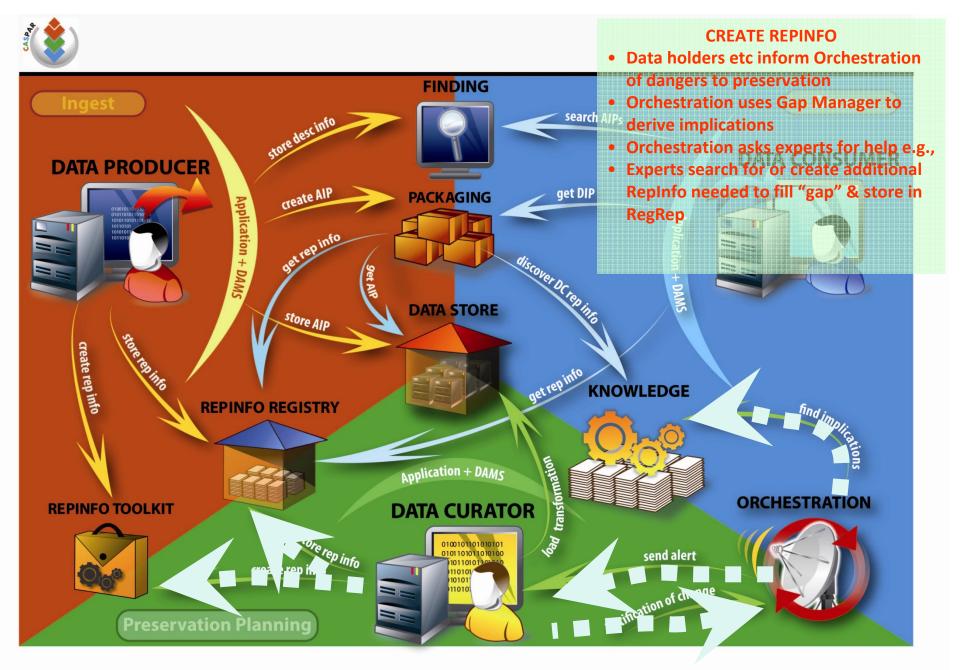




http://www.youtube.com/watch?v=Yun9hkPPF9M



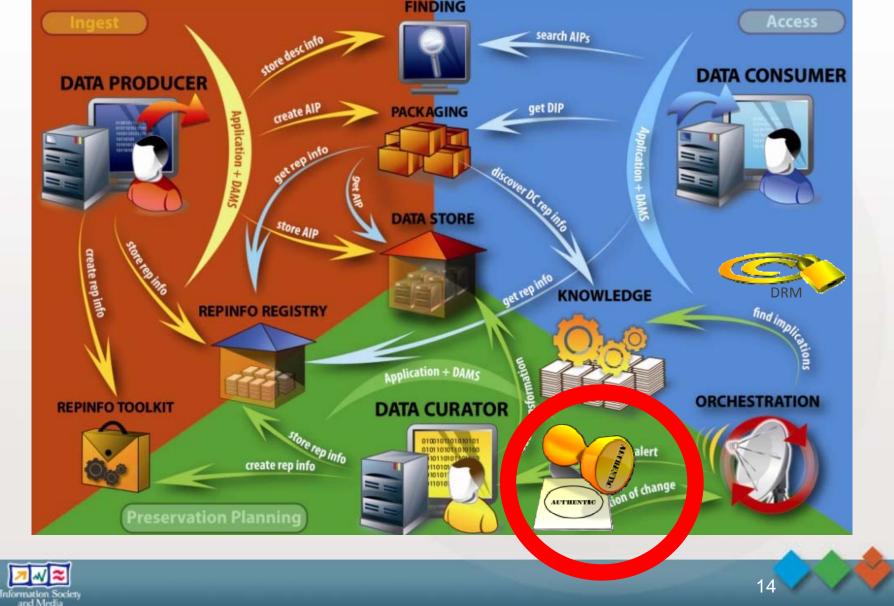


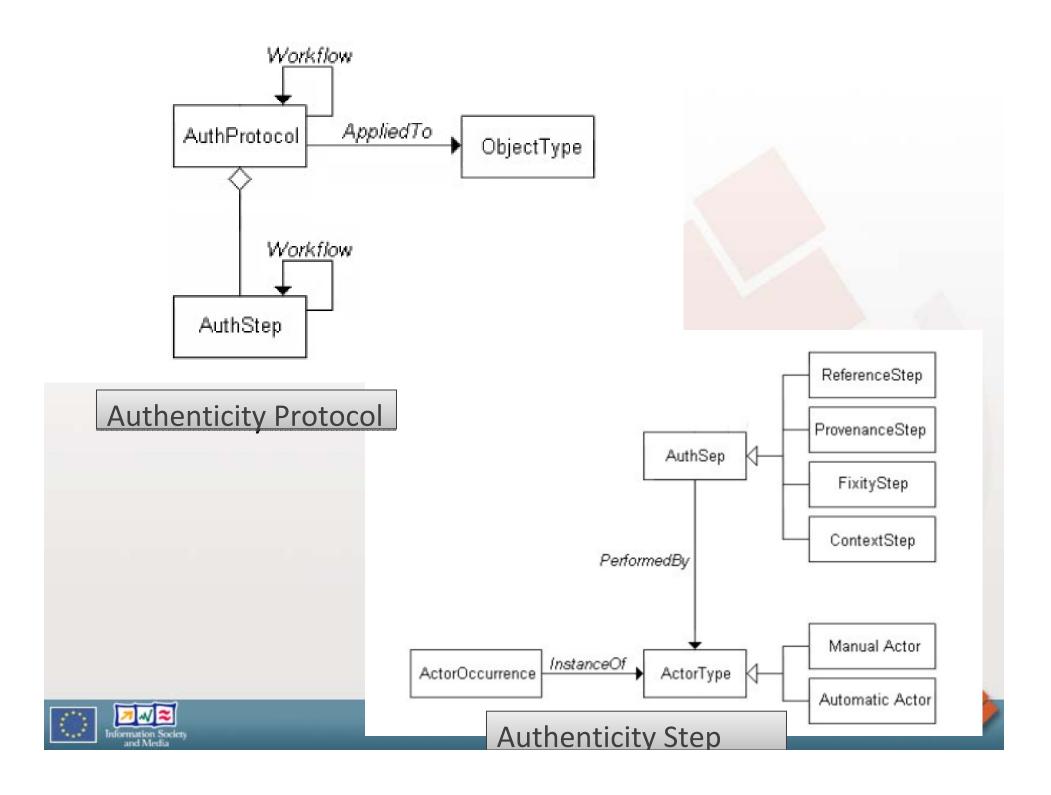


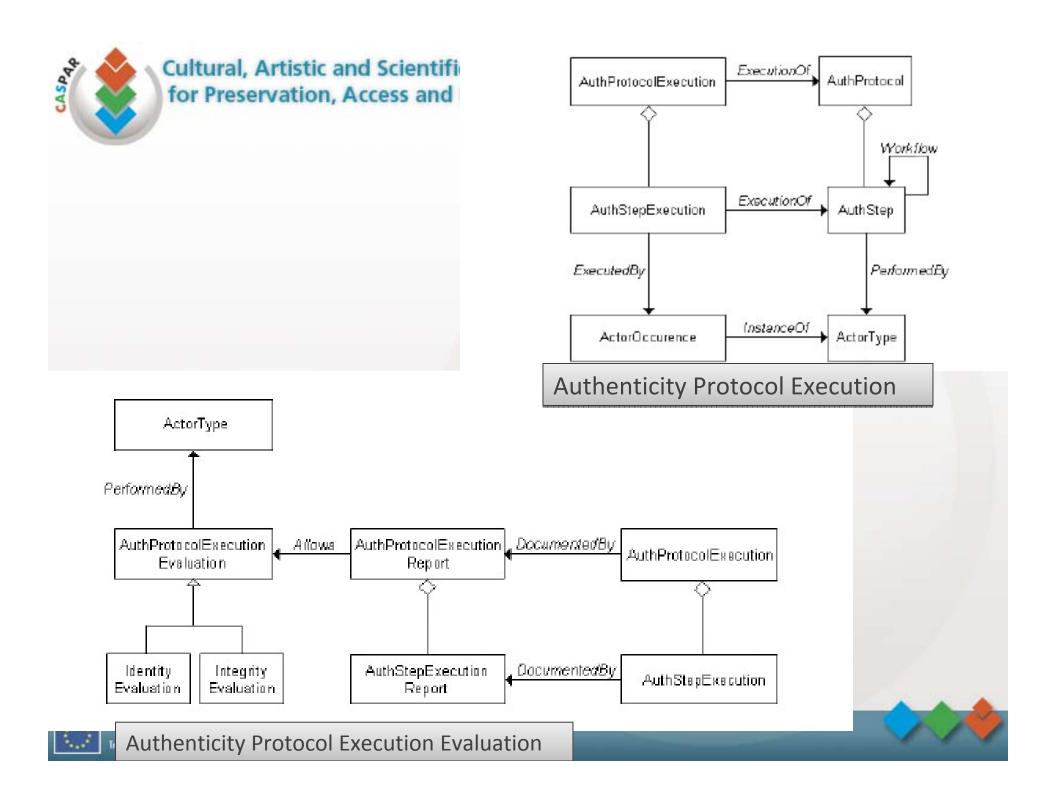




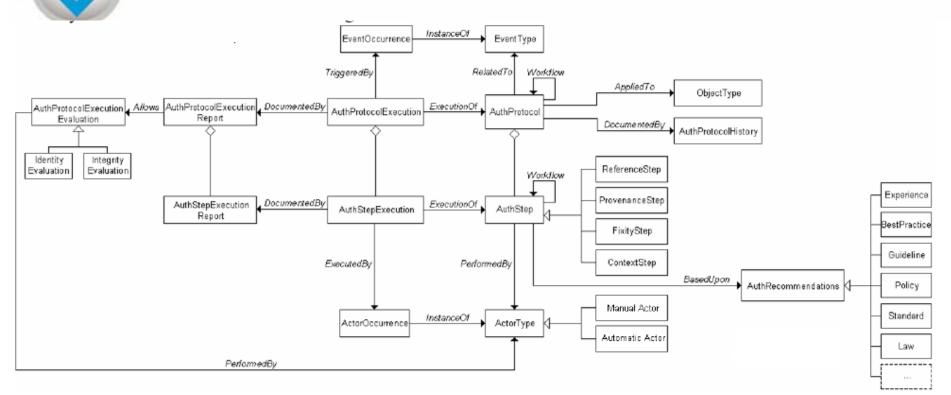
CASPAP

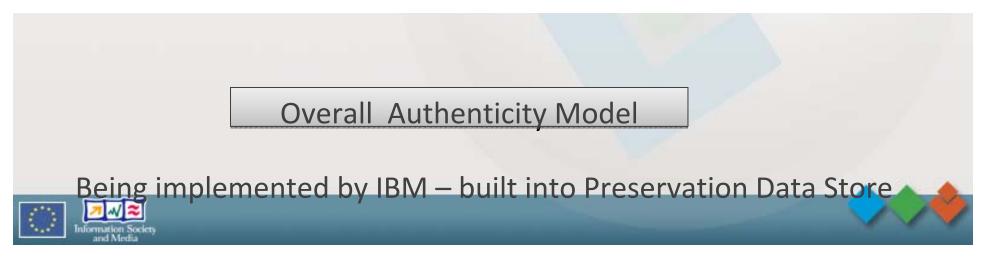














Transformation Protocol

Transformation of received data files into IIWG format

Recommendation (Policy)

"For the received data file to be deemed as sufficient quality to support data analysis it must have been successfully transformed into standard IIWG data format, the use of processing software must be recorded"

• Steps will capture

- Details of transformation used
 - Software details including name, version, source
 - Time/date of process
 - System details
 - Details of person responsible
 - Reason for believing this software is reliable
- Details of Transformational Information Properties checked
 - Information Property descriptions
 - Values checked
 - Details of how checked
 - Details of who was responsible for the check



Browse the Authenticity Catalogue elect a Project: Select elect a Project: Select totatactName ewinstance1 21-08-2009 12:08:08 Matt Dunckley View rotocol Step PDI Value Confidence ngestion Source of dataset Station ID tocation BUDAPEST - HUNGARY Geographic Latitude (WGS-84) 47.00*N Geographic Latitude (WGS-84) 47.00*N Geographic Latitude (WGS-84) 49.00*E Magnetic Latitude (UGRF-10(2005)) BUDAPEST Station Name BUDAPEST Source ContactName Instrument Name	thenticity P	rojects	Reports Accou	int Help	Log Out		
elect a Project: Select View istance Name Instance Created Project Owner Instance Status View View wiew wistance1 21-08-2009 12:08:54 Matt Dunckley View rotocol Step PDI Value Confidence I47 Captured by: Matt Dunckley on 09-09-2009 15:09:36 Station ID Station Location BUDAPEST - HUNGARY Geographic Latitude (WGS-84) 19.00°E Magnetic Latitude (WGS-84) 19.00°E Magnetic Latitude (GGF-10(2005))) BUDAPEST Station Name Instrument Instrument							
Instance Name Instance Created Project Owner Instance Status View Capture ewInstance1 21-08-2009 12:08:08 Matt Dunckley View ewInstance2 21-08-2009 12:08:54 Matt Dunckley View rotocol Step PDI Value Confidence Ingestion Source of dataset Station ID Istance Status View Confidence Station Location Location Latitude (WGS-84) 47.00°N Geographic Latitude (WGS-84) 19.00°E Magnetic Latitude (IGRF-10(2005)) BUDAPEST Station Name Source ContactName Instrument	» Brows	e the Auth	nenticity Catalo	gue			
ewInstance1 21-08-2009 12:08:08 Matt Dunckley View ewInstance2 21-08-2009 12:08:54 Matt Dunckley View rotocol Step PDI Value Confidence ngestion Source of dataset Station ID 147 Captured by: Matt Dunckley on 09-09-2009 15:09:36 Station Location BUDAPEST - HUNGARY Geographic Latitude (WGS-84) 47.00°N Geographic Latitude (WGS-84) 19.00°E Magnetic Longitude (UGF-10(2005)) BUDAPEST Station Name Source ContactName Instrument	elect a Project	: Select	▼ View				
ewInstance2 21-08-2009 12:08:54 Matt Dunckley View rotocol Step PDI Value Confidence ngestion Source of dataset Station ID 147 009-09-2009 15:09:36 95 Station Location BUDAPEST - HUNGARY Geographic Latitude (WGS-84) 47.00°N Geographic Latitude (WGS-84) 19.00°E Magnetic Longitude (WGS-84) 19.00°E Magnetic Longitude (IGRF-10(2005)) 45.93°N 85 Station Name Magnetic Longitude (IGRF-10(2005)) BUDAPEST Magnetic Longitude (IGRF-10(2005)) N Source ContactName Source ContactName Instrument N	istance Name	Instan	ce Created	Project Owner	Instance Status	View Capture	
rotocol Step PDI Value Confidence ngestion Source of dataset Station ID Station Location Station Name BUDAPEST - HUNGARY Geographic Latitude (WGS-84) 17.00°N Geographic Latitude (WGS-84) 19.00°E Magnetic Latitude (IGRF-10(2005)) BUDAPEST Station Name Source ContactName Instrument N	ewInstance1	21-08-	2009 12:08:08	Matt Dunckley		View	
ngestion Source of dataset Station ID Station ID Station Location Station Station Name Source ContactName Instrument	ewInstance2	21-08-	2009 12:08:54	Matt Dunckley		<u>View</u>	
ngestion Station ID Captured by: Matt Dunckley on 09-09-2009 15:09:36 Station Station Location BUDAPEST - HUNGARY Geographic Location Latitude (WGS-84) 47.00°N Geographic Latitude (WGS-84) 19.00°E Magnetic Latitude (IGRF-10(2005)) 45.93°N Magnetic Longitude (IGRF-10(2005)) BUDAPEST Station Name Source ContactName Instrument	rotocol	Step	PDI	Value		Confidence	
Captured by: Matt Dunckley on 09-09-2009 15:09:36 Station Location BUDAPEST - HUNGARY Geographic Latitude (WGS-84) 47.00°N Geographic Longitude (WGS-84) 19.00°E Magnetic Latitude (IGRF-10(2005)) 45.93°N Magnetic Longitude (IGRF-10(2005)) BUDAPEST Station Name Source ContactName Instrument	ngestion		Station ID			95	
Latitude (WGS-84) 47.00°N Geographic Location Longitude (WGS-84) 19.00°E Magnetic Latitude (IGRF-10(2005) 45.93°N Magnetic Longitude (IGRF-10(2005)) BUDAPEST Station Name Source ContactName Instrument		dataset					
Magnetic Longitude (IGRF-10(2005)) BUDAPEST Station Name Source ContactName Instrument				Latitude (WGS-84 Longitude (WGS-8) 47.00°N Geographic 4) 19.00°E Magnetic	*	
ContactName Instrument			Station Name	Magnetic Longitud BUDAPEST			N



20



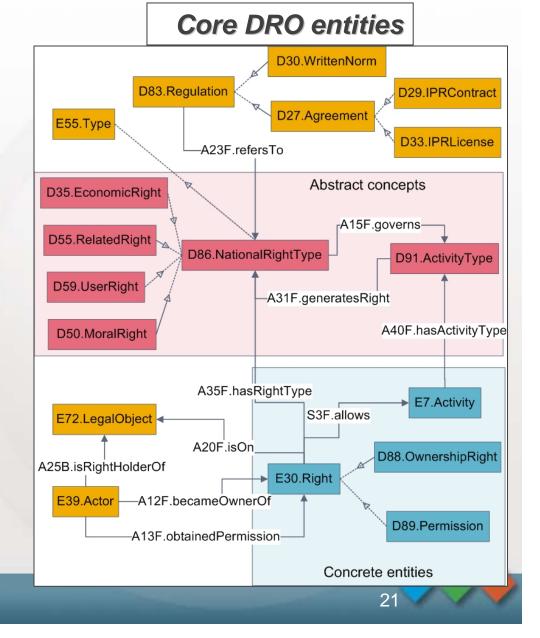
CASPAP

Cultural, Artistic and Scientific knowledge for Preservation, Access and Determinal Rights Ontology

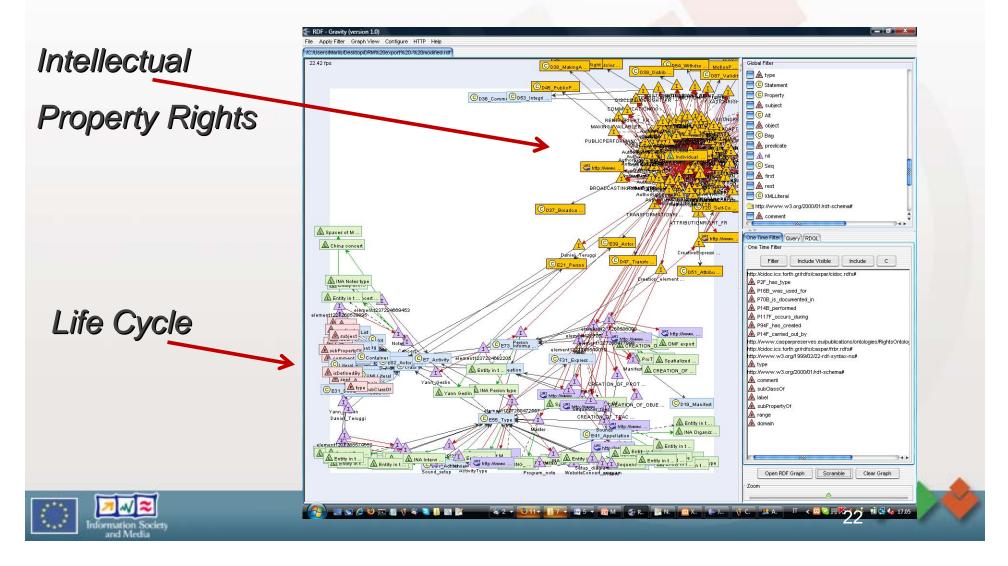
A Domain ontology

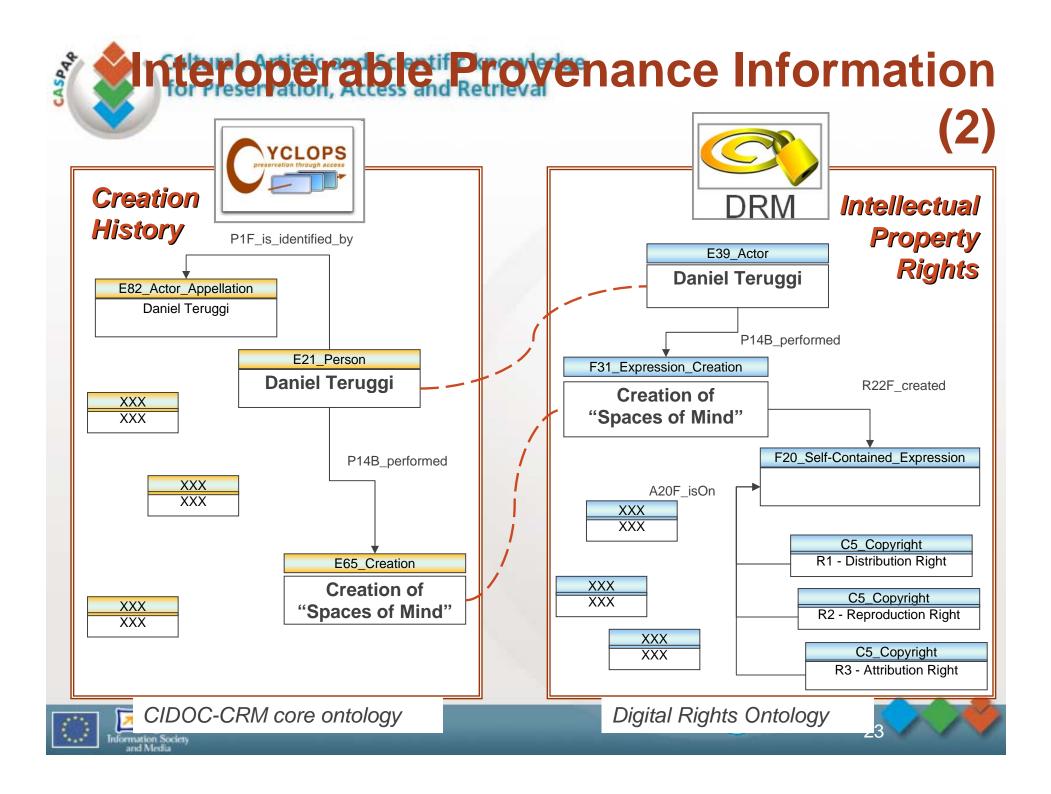
CASPAP

- About Intellectual Property Rights
- To describe all aspects that are relevant to evaluate rights
 - Attributes of rights
 - Validity in time and space
 - Laws and Agreements
 - People
 - Actions
 - Licenses
 - Constraints and Conditions
- Harmonized with CIDOC CRM and FRBRoo



Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Interoperable Provenance Information (1)







Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Insight: stakeholders

Funding/policy

- National Funding
- organisations
- European funding
- Corporate funding

Research

- Research institutes (non-profit)
- Universities
- Academic libraries

Data management (preservation)
Data centres (profit / nonprofit)
Libraries
Archives Publishing

- General (cross-community)
- publishers
- Specific (community) publishers





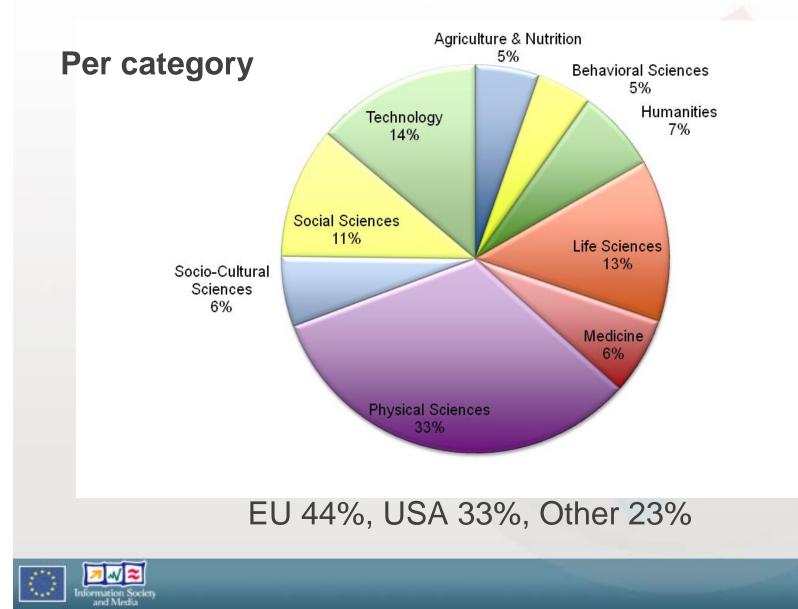
Funding/policy
< responses

Research 1397 responses

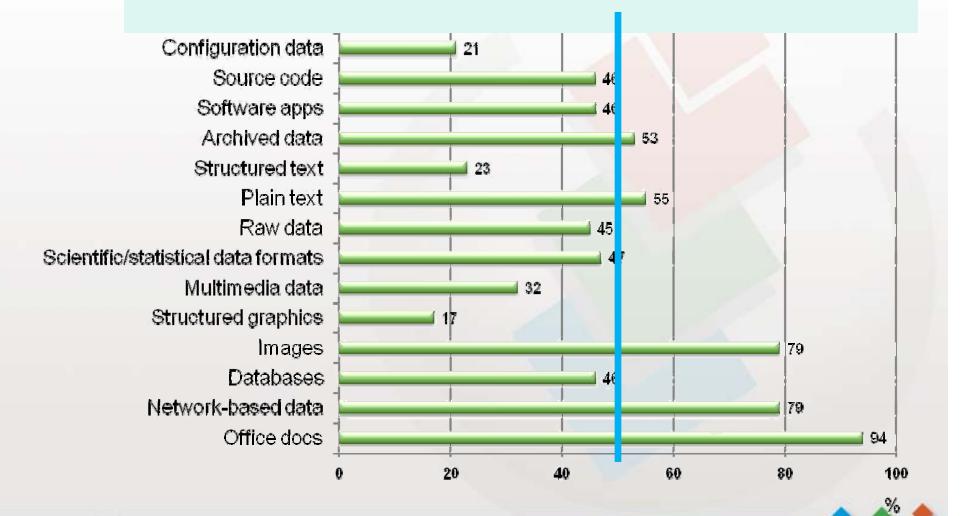
Data management (preservation) 273 responses Publishing 186 responses













Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Sharing of data (R) How open is your data?

I do not share my data and I do not want to share it in the 6,3% future. 16,2% I do not share my data, but I would like to 10,8% My data could be made available with appropriate changes 3,5% My data is available for a fee. 15.8% Access to my data is temporarily restricted. 58,7% My data is openly available for my research group 9,9% My data is openly available for my research discipline. 24,3% My data is openly available for everyone.

0

0,1

0,2

0,4

0,3

0,5

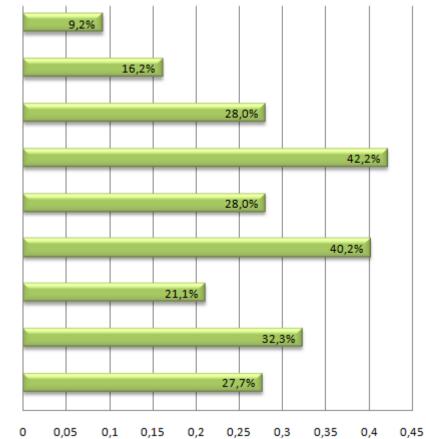
0,6

0,7

Information Society and Media

Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Sharing of data (R)

Which constrains do you see in making data open?



Other (please specify) No problems foreseen Lack of financial resources Misuse of data Lack of technical infrastructure Legal issues Restricted access to data archive Incompatible data types Fear to lose scientific edge



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

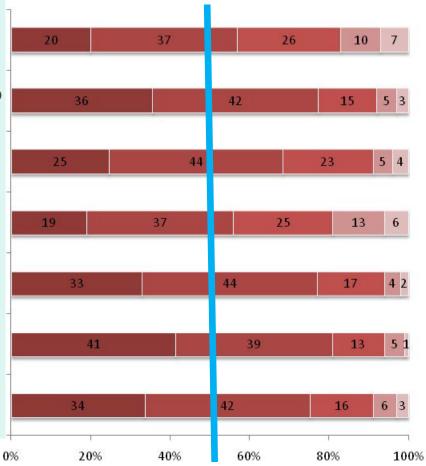
Loss of ability to identify the location of data

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



Very important

Slightly Important

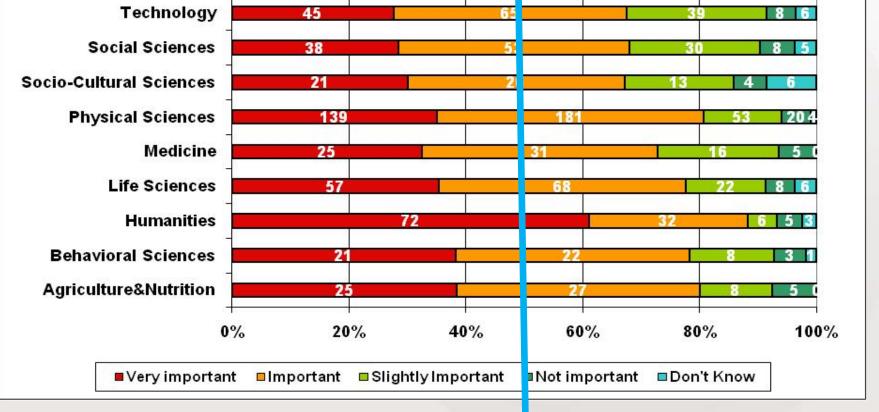
Not important Don't Know







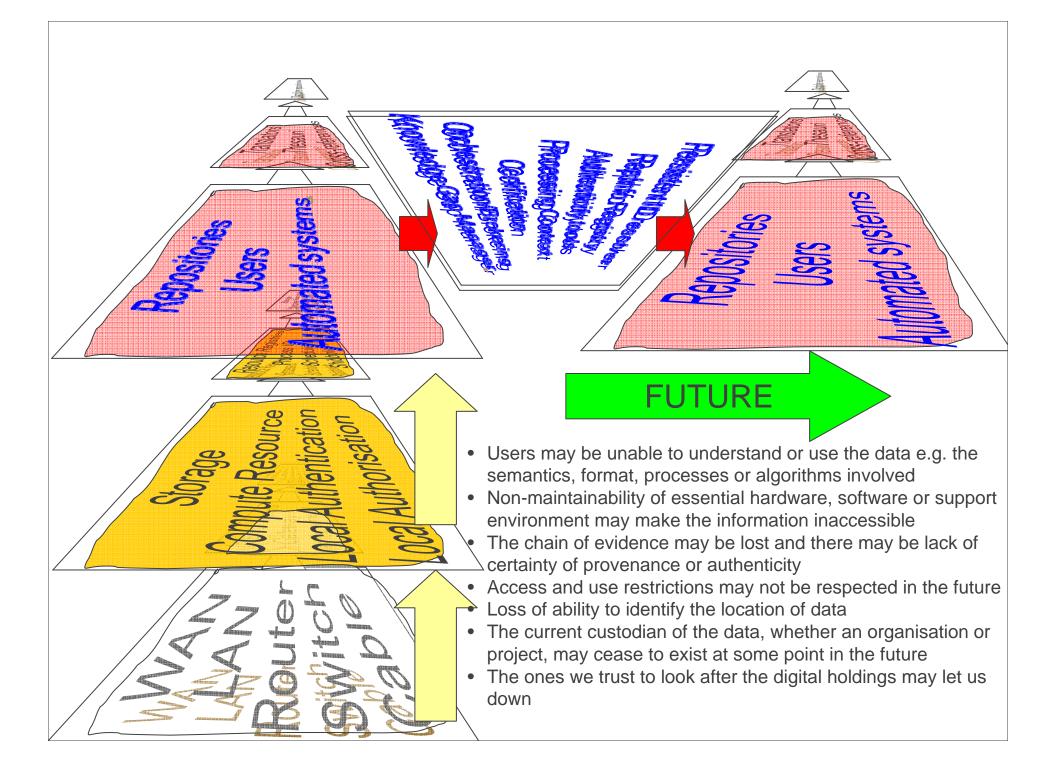








Threat	Requirement for solution
Users may be unable to understand or use the data e.g. the semantics, format, processes or algorithms involved	
Non-maintainability of essential hardware, software or support environment may make the information inaccessible	
The chain of evidence may be lost and there 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	



• CASPAR:

- www.casparpreserves.eu
- http://www.casparpreserves.eu/training/advanced-digitalpreservation-training loctures/ http://developers.casparpreserves.eu:8080
- PARSE.Insight:
 - www.parse-insi http://sourceforge.net/projects/digitalpreserve
- Alliance for Permanent Access:
 - www.alliancepermanentaccess.eu
- Digital Curation Centre:
 - www.dcc.ac.uk
- YouTube cartoon:
 - www.youtube.com/watch?v=Yun9hkPPF9M
- OAIS:
 - http://public.ccsds.org/publications/archive/650x0b1.pdf
 - http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Overview.aspx





END

← Alliance for Permanent Access →



The Consultative Committee for Space Data Systems









Communities aggregated to:

- Agriculture & Nutrition
- Behavioural Sciences
- Humanities
- Life Sciences
- Medicine
- Social Sciences
- Physical Sciences
- Socio-Cultural Sciences
- Technology

Based on KNAW classification (Royal Netherlands Academy of Arts and Sciences)





- 1. It is unique.
- 2. It potentially has economic value.
- 3. It may stimulate inter-disciplinary collaborations.
- 4. It allows for re-analysis of existing data.
- 5. It may serve validation purposes in the future.
- 6. It will stimulate the advancement of science (new research can build on existing knowledge).
- 7. If research is publicly funded, the results should become public property and therefore properly preserved.

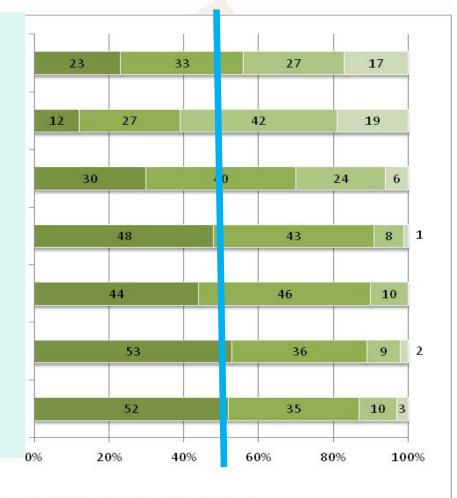




Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Reasons for preservation (R)

It is unique

- It potentially has economic value
- It may stimulate inter-disciplinary collaborations.
- It allows for re-analysis of existing data.
- It may serve validation purposes in the future.
- It will stimulate the advancement of science.
- If research is publicly funded, the results should become public property and therefore properly preserved.



🔳 Very Important 📲 Important 📲 Slightly Important 📲 Not Important



Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Reasons for preservation (DM)

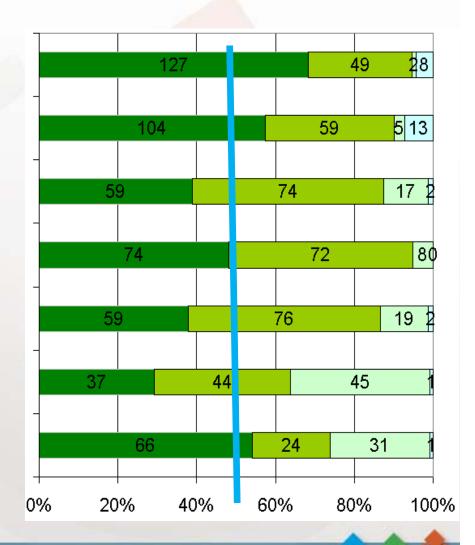
It is unique

It potentially has economic value

It may stimulate inter-disciplinary collaborations.

It allows for re-analysis of existing data.

- It may serve validation purposes in the future.
- It will stimulate the advancement of science.
- If research is publicly funded, the results should become public property and therefore properly preserved.





Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Reasons for preservation (P)

It is unique

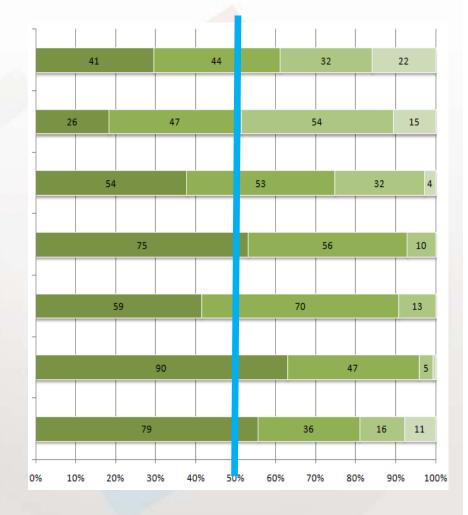
It potentially has economic value

It may stimulate inter-disciplinary collaborations.

It allows for re-analysis of existing data.

It may serve validation purposes in the future.

- It will stimulate the advancement of science.
- If research is publicly funded, the results should become public property and therefore properly preserved.





Very Important

Important Slightly

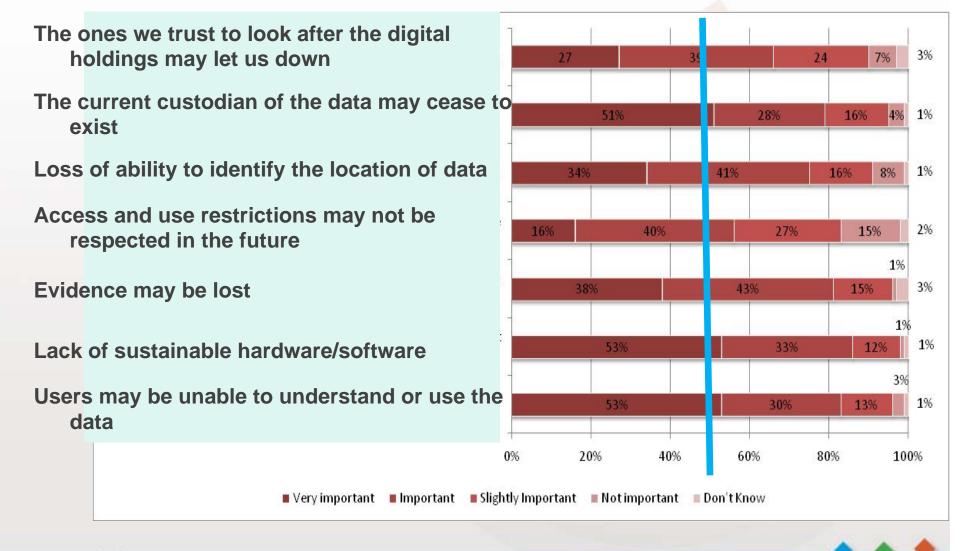
Slightly Important
Not Important



- 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.



Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Threats to preservation (DM)





CASPAD

Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Threats to preservation (P)

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

The current custodian of the data may cease to exist

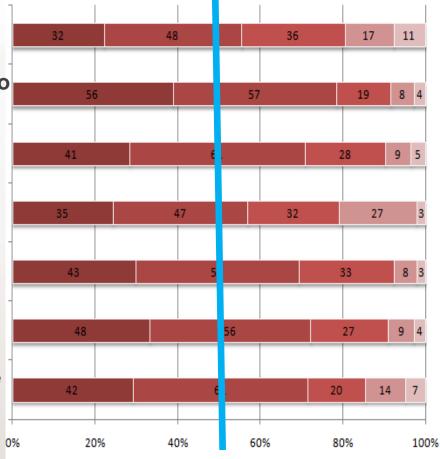
Loss of ability to identify the location of data

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





Very Important
Important

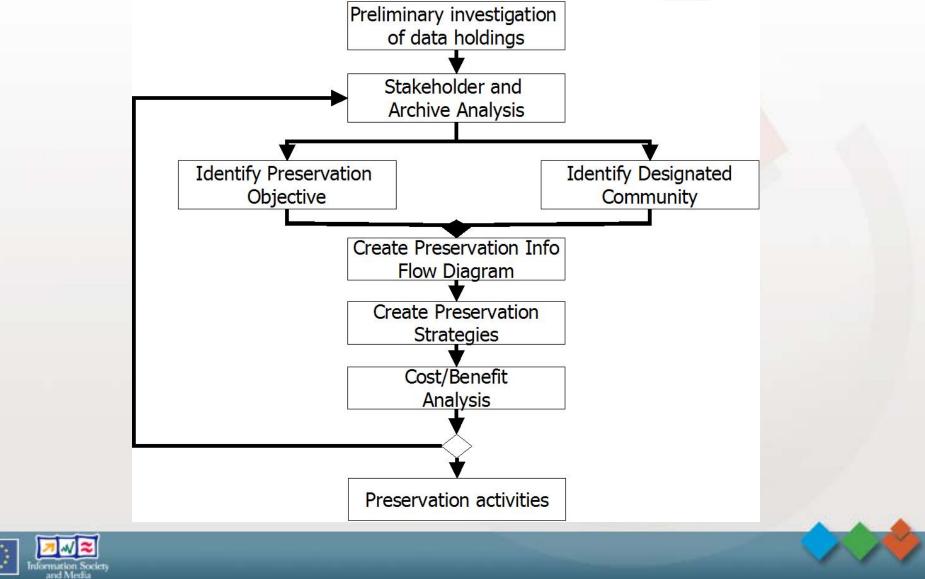
Slightly Important

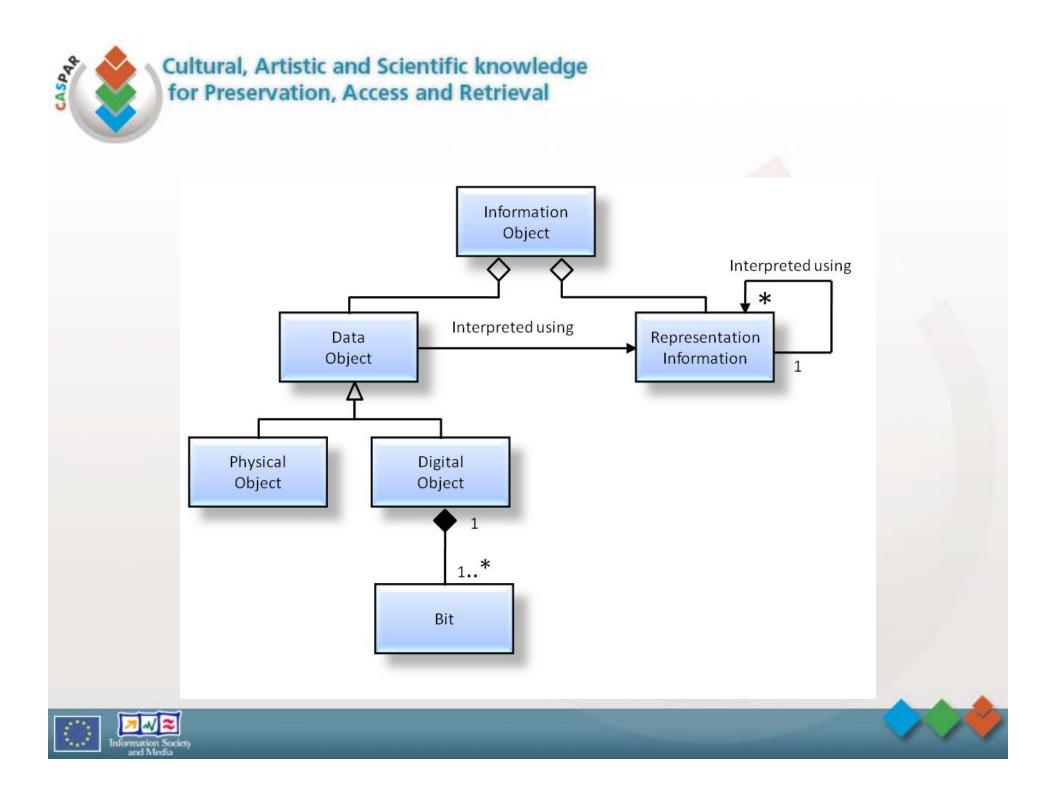
Not Important Don't Know

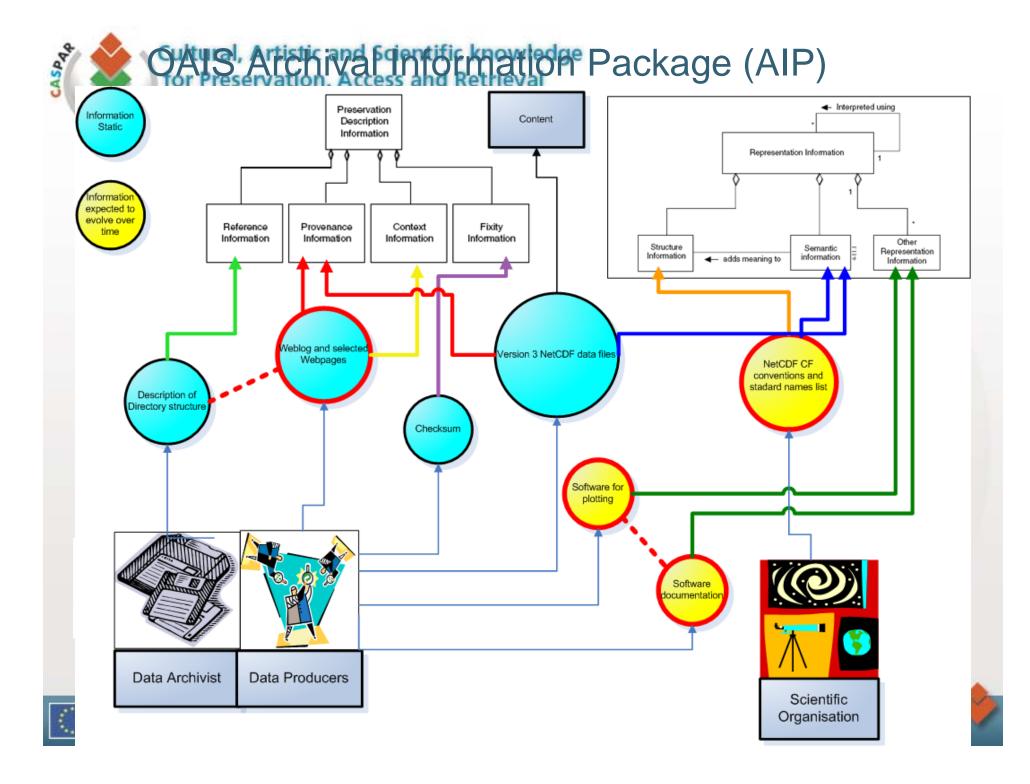


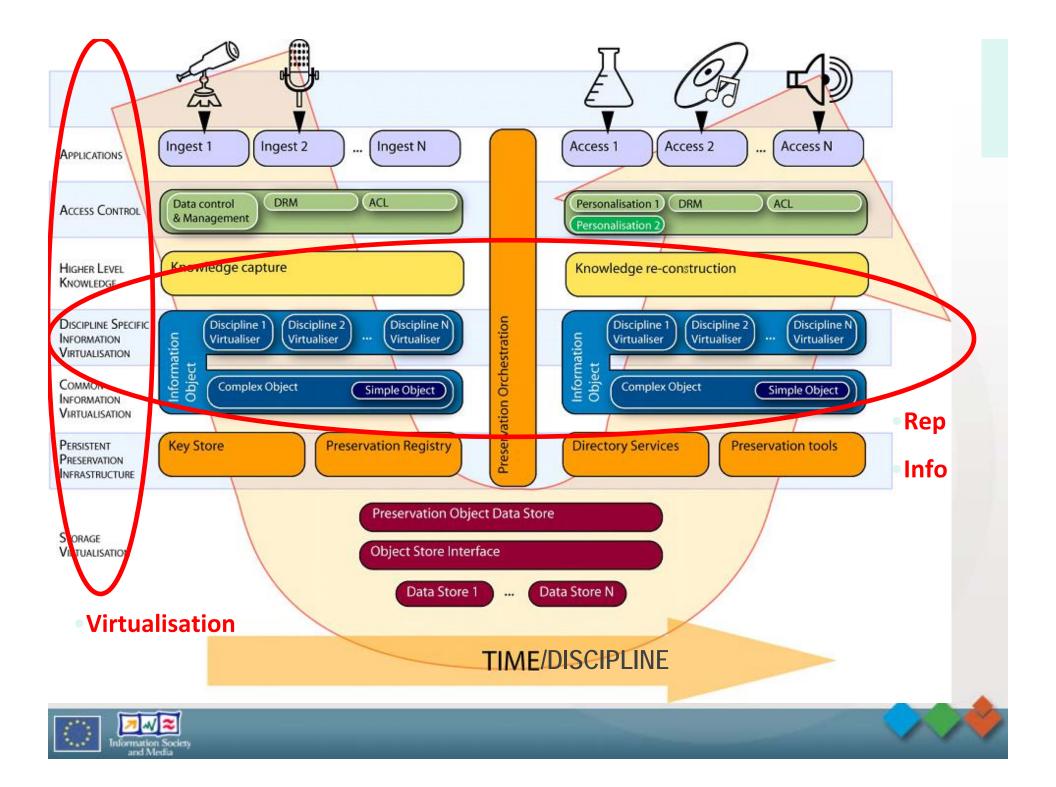
CASPAR











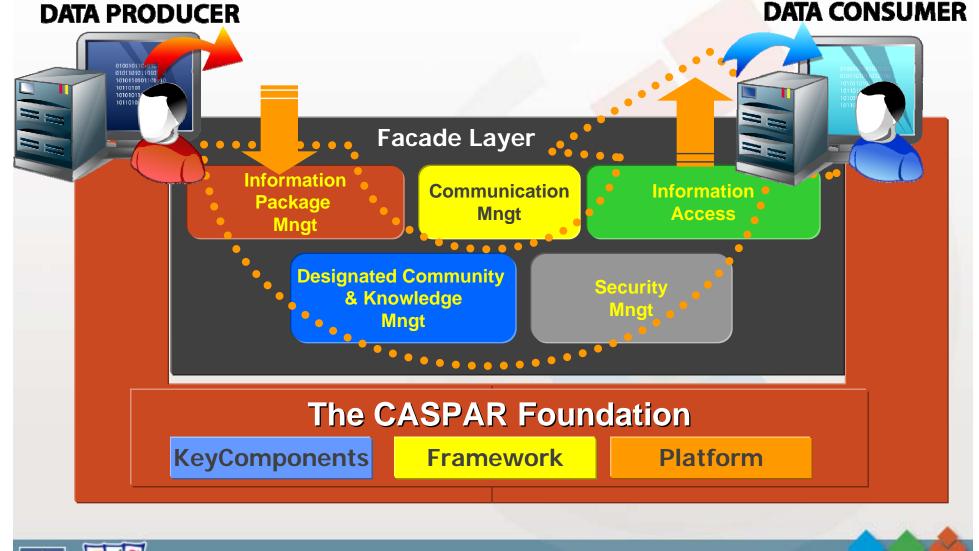


EU FP6 Integrated Project

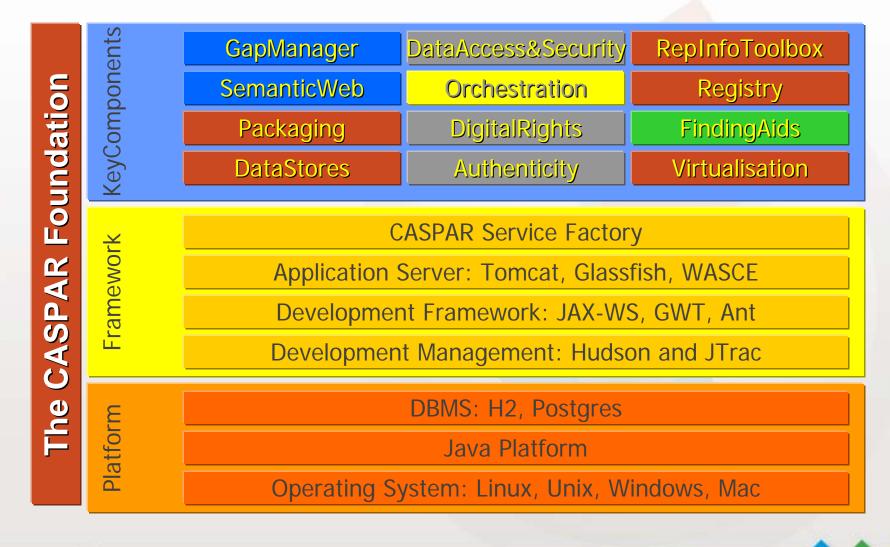
Total spend approx. 16MEuro (8.8 MEuro from EU)















Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Preservation Issues 1...

- 1. How To guarantee a digital information may be accessed and understood in the future
- 2. How To guarantee a proper information package management within and OAIS Archive
- 3. How To guarantee long-time preservation maintenance of any information package
- 4. How To guarantee retrieval of Archival Information



Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Preservation Issues 2...

- 4. How To guarantee intellegibility within heterogeneous Designated Communities and their digital information
- 5. How To guarantee preservation actors are informed about change events
- 7. How To guarantee an adequate security access with the proper rights to any resource and functionality within an OAIS Archive
- 8. How To guarantee an adequate integrity and identity for any Archival Information





 To guarantee a digital information may be accessed and understood in the future, you need an adequate OAIS **Representation Information**









Virtualisation





REG

Registry



• To guarantee a proper information package management within and OAIS Archive, you need to create an adequate OAIS Information Package









• To guarantee long-time preservation maintenance of any information package, you need an implementation of OAIS Archival Storage



Preservation DataStores

PDS







 To guarantee retrieval of Archival Information, you need an OAIS Finding Aids









 To guarantee intellegibility within heterogeneous Designated Communities and their digital information, you need to manage Designated Community Profiles and their Knowledge Base KM







 To guarantee preservation actors are informed about change events, you need an adequate management of message exchange









 To guarantee an adequate security access with the proper rights to any resource and functionality within an OAIS Archive, you need a Security and DRM Management







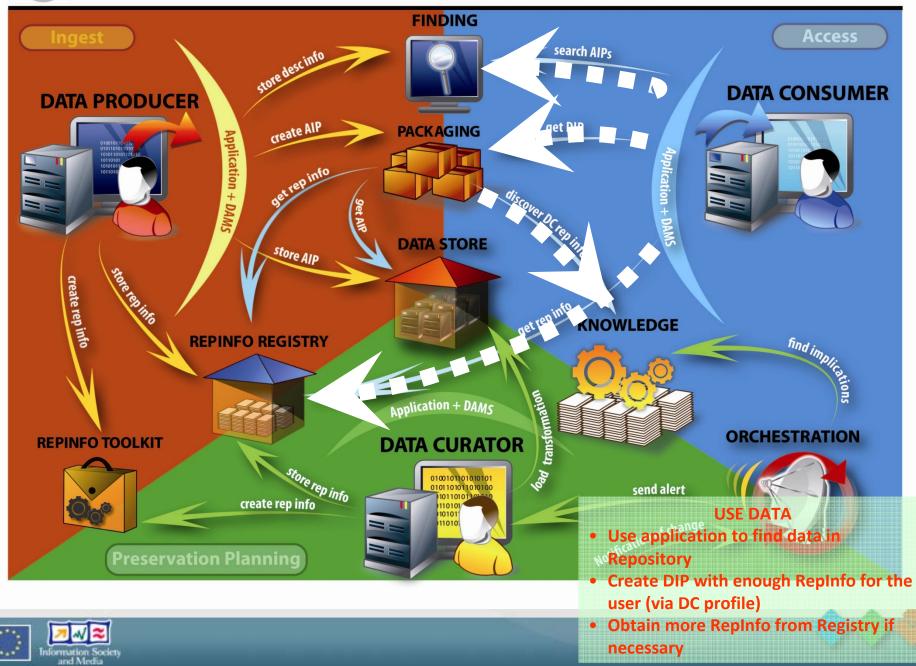
• To guarantee an adequate integrity and identity for any Archival Information, you need an Authenticity Tool



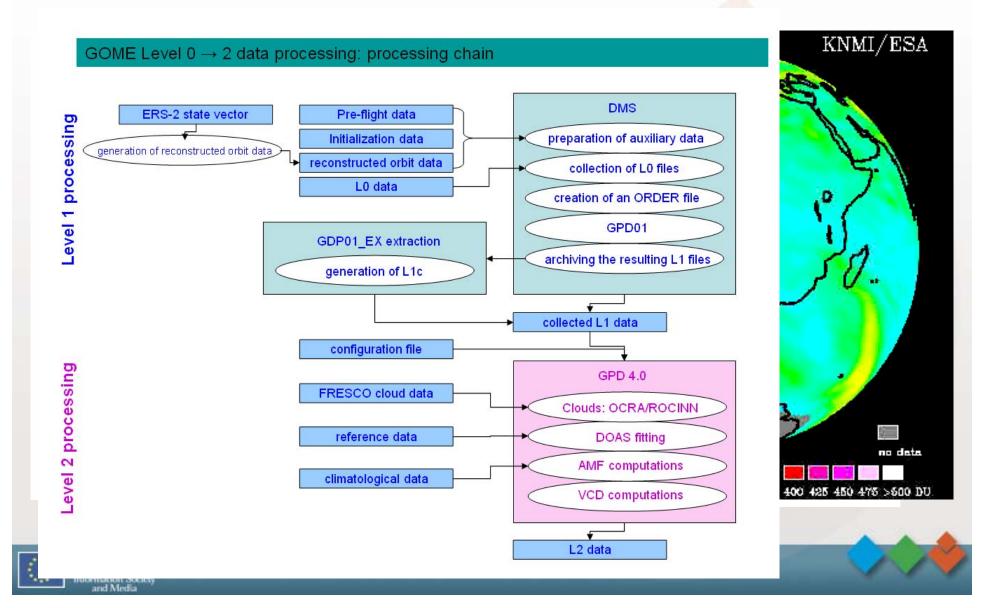




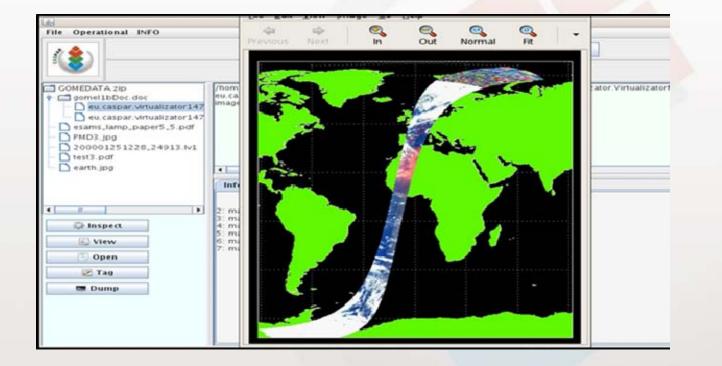










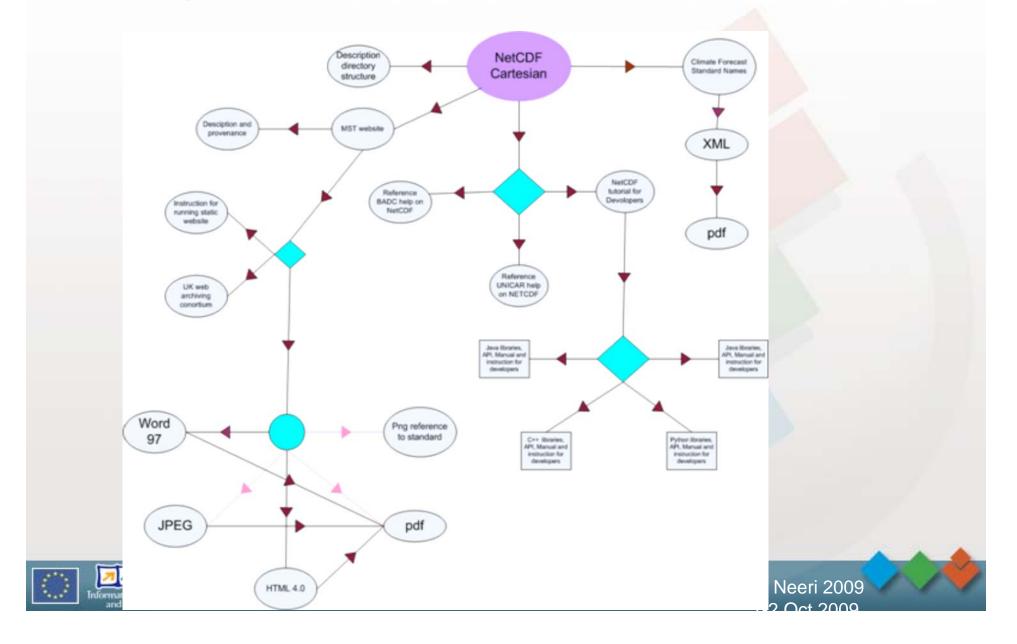






Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Representation Information Network

CASPAP





RFADMF.txt

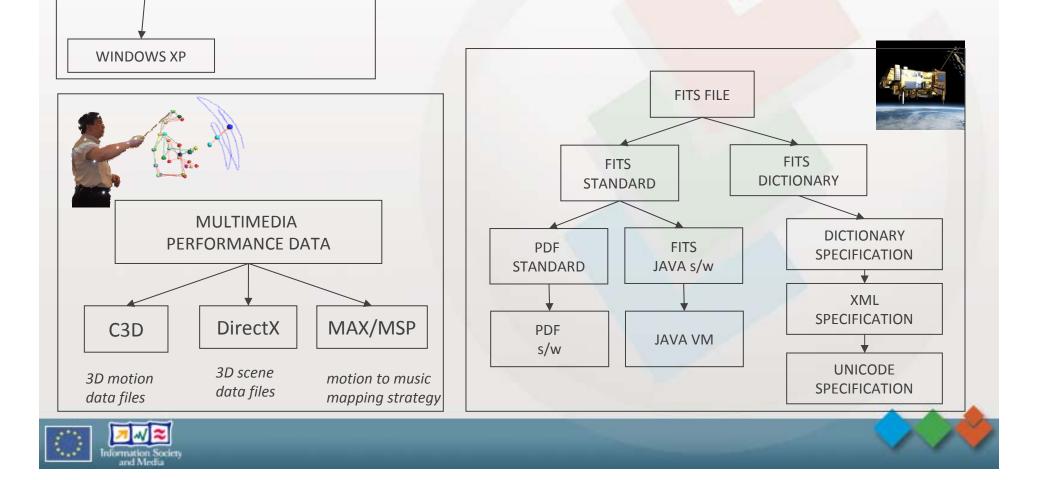
TEXT EDITOR

ENGLISH

LANGUAGE

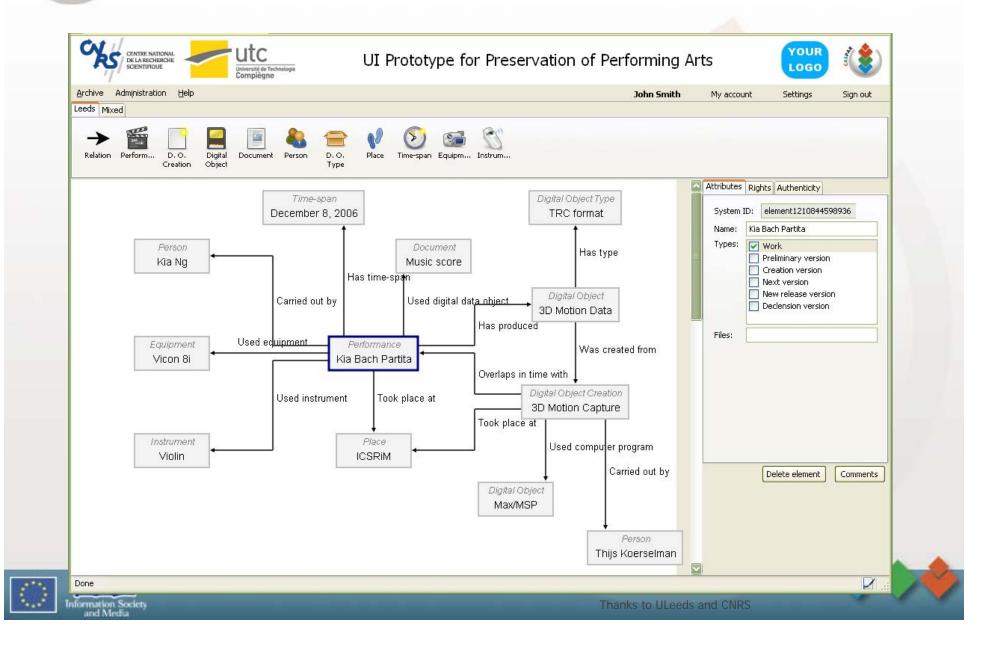
Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

Modules and Dependencies: defining the Designated Community

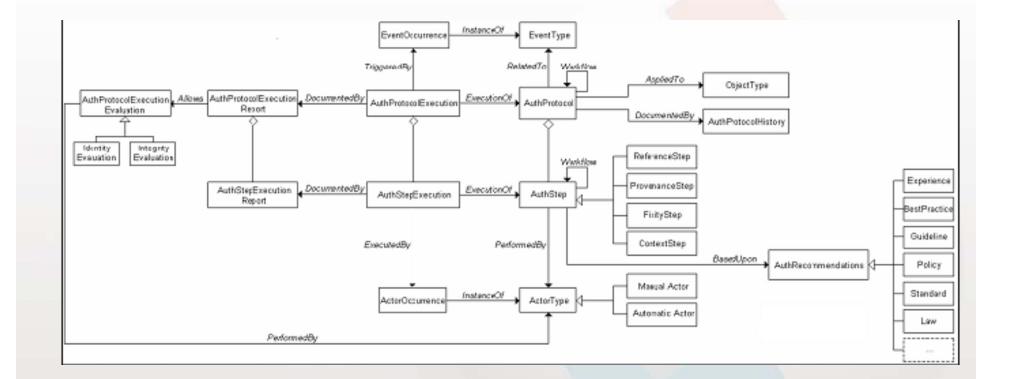


Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Provenance: Performing Arts

CASPAD



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





CASPAP

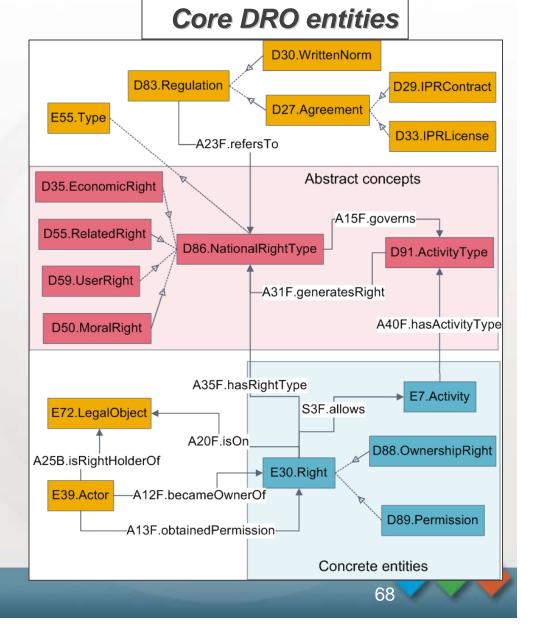


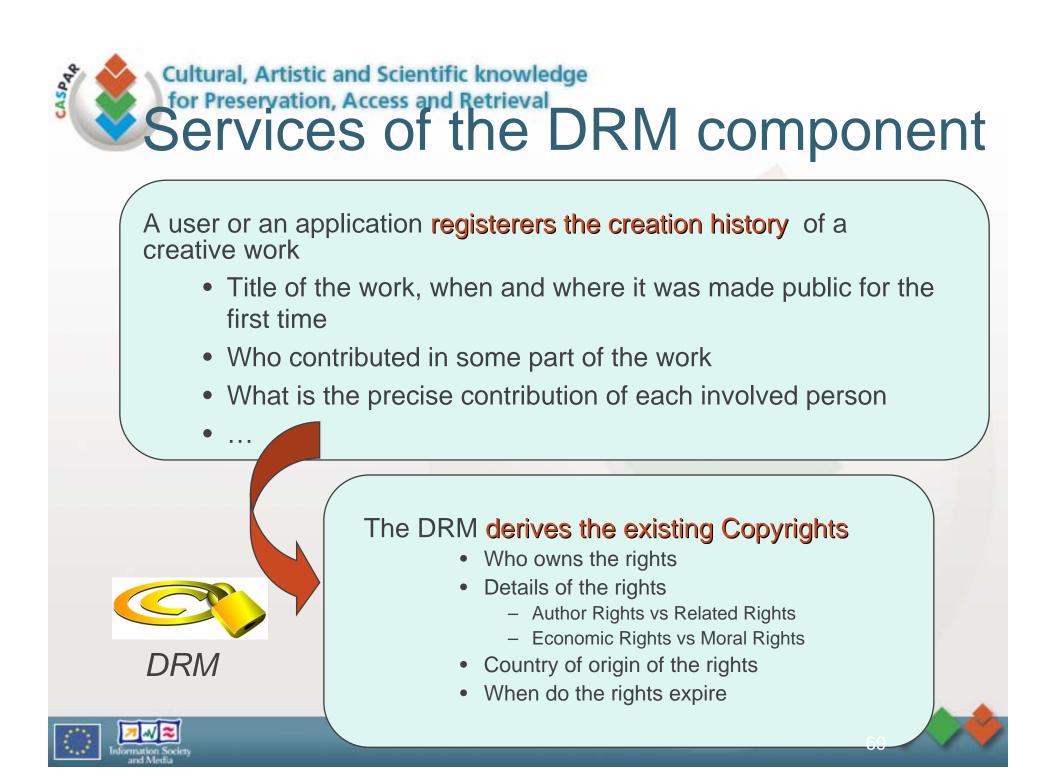
Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval DIGITAL RIGHTS ONTOLOGY

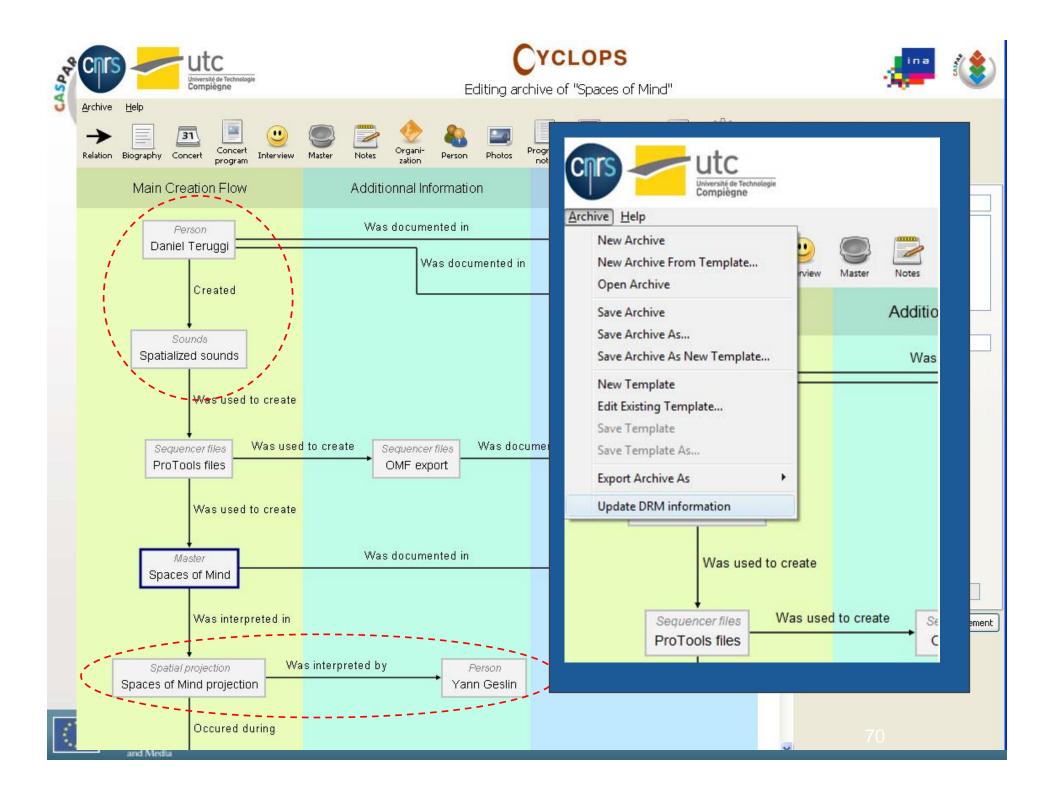
A Domain ontology

CASPAD

- About Intellectual Property Rights
- To describe all aspects that are relevant to evaluate rights
 - Attributes of rights
 - Validity in time and space
 - Laws and Agreements
 - People
 - Actions
 - Licenses
 - Constraints and Conditions
- Harmonized with CIDOC CRM and FRBRoo







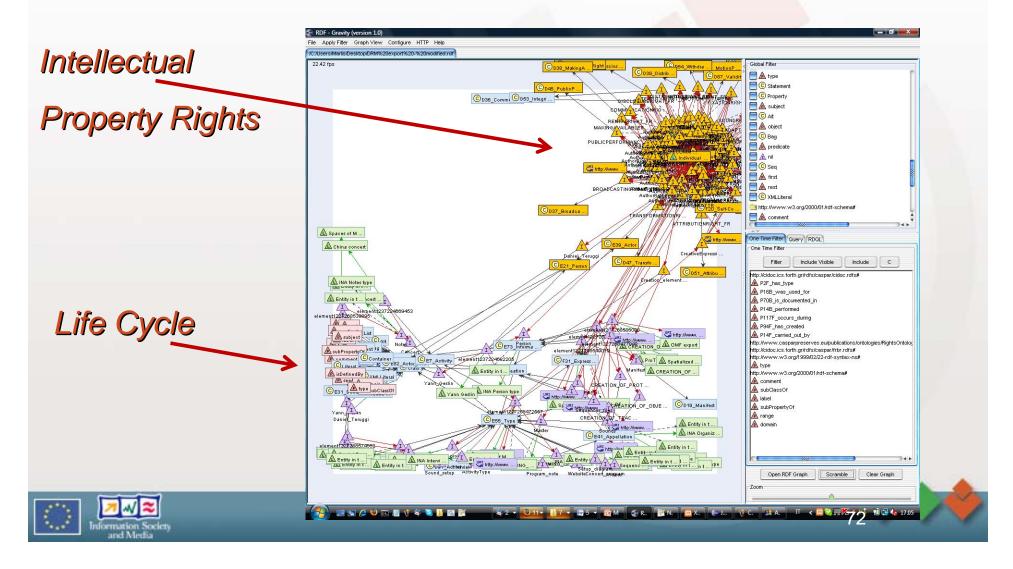


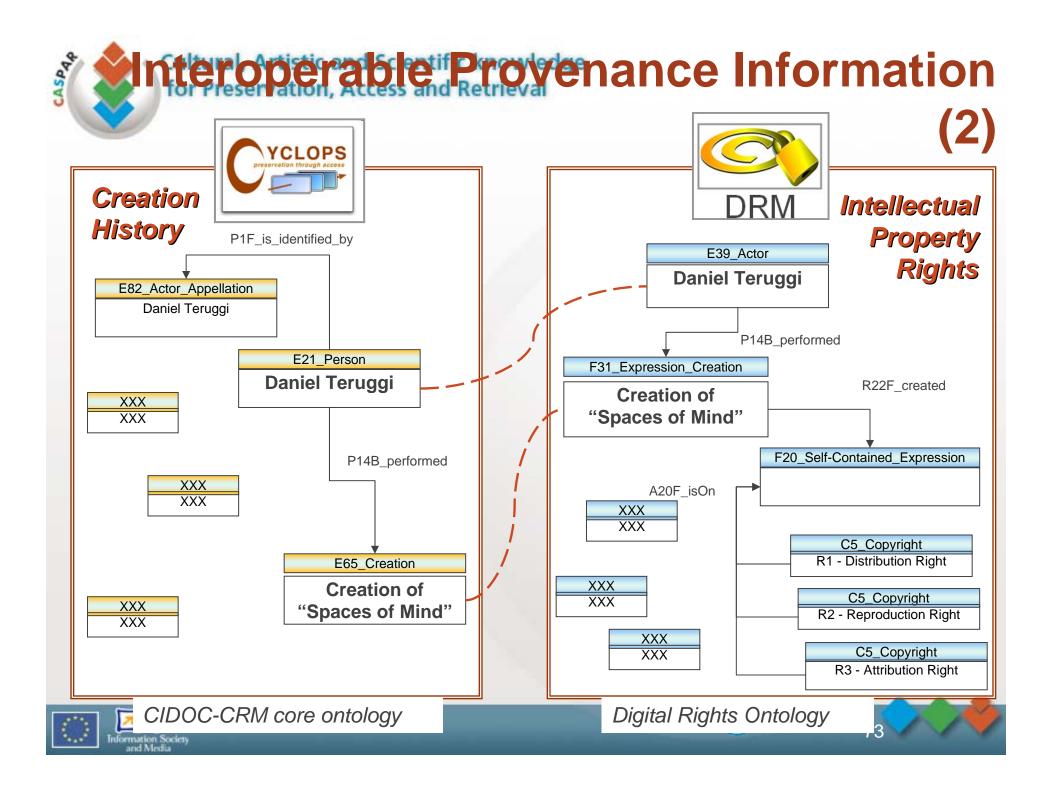
Cultural Artistic and Scientific knowledge

😉 Source of: file:///C:/Documents%20and%20Settings/Marlis/Desktop/3.rdf - Mozilla Firefox File Edit View Help <rdf:RDF xmlns:cidoc="http://cidoc.ics.forth.gr/rdfs/caspar/cidoc.rdfs#" xmlns:cidoc-digital="http://cidoc.ics.forth.gr/rdfs/caspar/cidoc digital.rdfs#" xmlns:rights-onto="http://www.casparpreserves.eu/publications/ontologies/RightsOntology/rights-ontology-v0 5 5 xmlns:frbr="http://cidoc.ics.forth.gr/rdfs/caspar/frbr.rdfs#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://www.w3.org/2002/07/ow1#" xmlns:drm-data="http://eu.casparpreserves/resource/DRM/data#" > <rdf:Description rdf:about="http://eu.casparpreserves/resource/DRM/data#ATTRIBUTIONRIGHT FR"> <rdfs:comment xml:lang="en">Right as defined in FR on 18-03-2009</rdfs:comment> <rdf:type rdf:resource="http://www.casparpreserves.eu/publications/ontologies/RightsOntology/rights-ontology-v </rdf:Description> <rdf:Description rdf:about="http://eu.casparpreserves/resource/DRM/data#TRANSFORMATIONRIGHT FR"> <rdfs:comment xml:lang="en">Right as defined in FR on 18-03-2009</rdfs:comment> <rdf:type rdf:resource="http://www.casparpreserves.eu/publications/ontologies/RightsOntology/rights-ontology-v </rdf:Description> <rdf:Description rdf:about="http://eu.casparpreserves/resource/DRM/data#RENTALRIGHT FR"> <rdfs:comment xml:lang="en">Right as defined in FR on 18-03-2009</rdfs:comment> <rdf:type rdf:resource="http://www.casparpreserves.eu/publications/ontologies/RightsOntology/rights-ontology-v </rdf:Description> <rdf:Description rdf:about="http://eu.casparpreserves/resource/DRM/data#spaces of mind"> <rdf:type rdf:resource="http://cidoc.ics.forth.gr/rdfs/caspar/cidoc.rdfs#E35 Title"/> </rdf:Description> <rdf:Description rdf:about="http://eu.casparpreserves/resource/DRM/data#AuthorRight 1237377665607"> <rights-onto:A20F is0n rdf:resource="http://eu.casparpreserves/resource/DRM/data#CreativeExpression 1237377665; <rights-onto:A33F hasValidity rdf:resource="http://eu.casparpreserves/resource/DRM/data#Validity 06-03-2009 nu <rights-onto:A35F hasRightType rdf:resource="http://eu.casparpreserves/resource/DRM/data#TRANSFORMATIONRIGHT F <rdfs:comment xml:lang="en">Individual ownership right held by a person on a precise creative contribution to <rdf:type rdf:resource="http://cidoc.ics.forth.gr/rdfs/caspar/cidoc digital.rdfs#C5 Copyright"/> </rdf:Description> <rdf:Description rdf:about="http://eu.casparpreserves/resource/DRM/data#AuthorRight 1237377665737"> <rights-onto:A20F is0n rdf:resource="http://eu.casparpreserves/resource/DRM/data#CreativeExpression 1237377665 23 TH UAS VARIAN THE TEXTING

Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Interoperable Provenance Information (1)

CASPAD





Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval Scenario: Change in Copyright Law

- Goal
 - Demonstrate that the system withstands changes in Law

• Scenario:

- An amendment to a Copyright Law clarifies who can be considered a performer, hence eligible to related rights
 - Current IPR Law No. 92–597 of July 1, 1992, Art. 212-1 says:

"... persons who act, sing, deliver, declaim, play in or otherwise perform literary or artistic works, variety, circus or puppet acts."

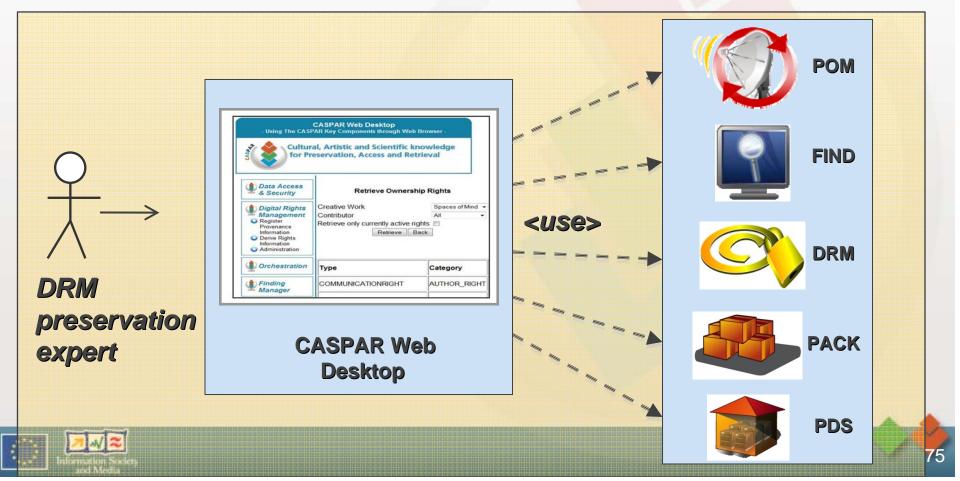
- A hypothetical amendment could extend the definition of performers including also people who project and interpret acousmatic sound files
- \rightarrow We have more rightholders





Update of PDI (Provenance)

Update Intellectual Property Rights after a Change in Law



Scenario Change in Copyright Law (2/2)

- 0. A notification is sent to the DRM Preservation Expert
 - Through the POM

Actions to do:

- 1. The rules for correctly deriving existing rights must be updated
 - Through the DRM Administration Interface
- 2. Rights information (PDI-Provenance) must be updated for all affected archive holdings
 - The affected AIPs are retrieved (FIND Key Component)
 - The rights are re-calculated (DRM Key Component)
- 3. The new Provenance is packaged in the AIPs and stored
 - Through the PACK and PDS Key Components



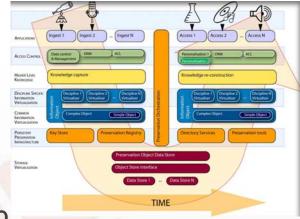


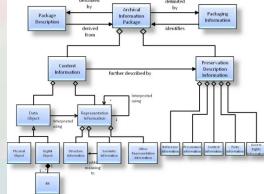
- Standard for certification in OAIS Roadmap
- Initial work produced TRAC
- Now an official CCSDS Working Group
- Open virtual meetings, notes and documents: <u>http://www.digitalrepositoryauditandcertification.org</u>
- Draft standard submitted to CCSDS/ISO to form the basis of an international audit and certification process





- Preservation
 - needs more than format information
 - needs more than Significant Properties
- Many layers of information (metadata)
- This extra information
 - must be created tools are needed
 - must itself be preserved
- Made easier by
 - being integrated with (e-)infrastructure
 - addressing users' concerns
- Must deal with all types of digital objects

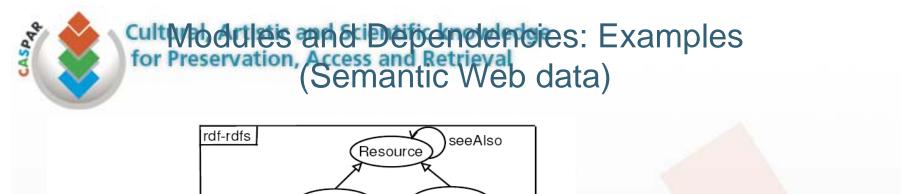


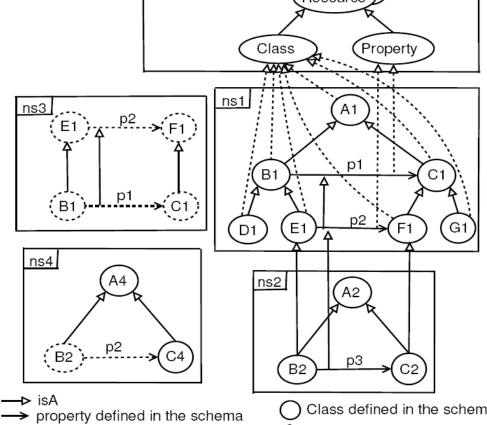


Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval From OAIS: OAIS conformance

- A conforming OAIS archive implementation shall support the model of information described in 2.2. The OAIS Reference Model does not define or require any particular method of implementation of these concepts.
- A conforming OAIS archive shall fulfill the responsibilities listed in 3.1. Subsection 4 provides examples of the mechanisms that may be used to discharge the responsibilities identified in 3.1. These mechanisms are not required for conformance. It is expected that a separate standard, as noted in section 1.5, will

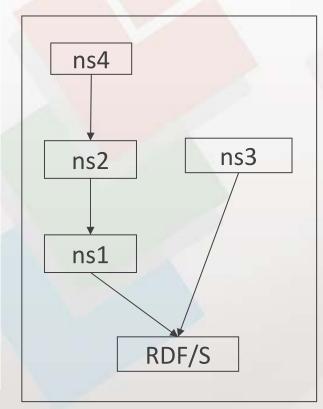
Produced on which accreditation and 79





- ---> property defined in other schemas --- > instanceOf
- Class defined in the schema ()Class defined in other schemas

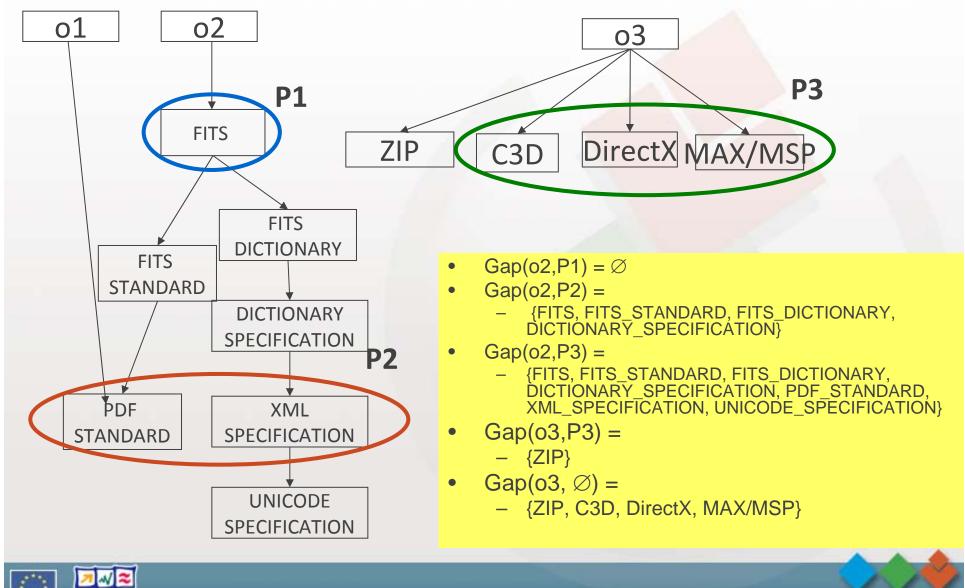
modules and dependencies







Cultural, Artistic and Scientific knowledge Scenariotionterligibility aware Packaging



ASD.	Threat	Requirement for solution
0	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	



