



**APA**

ALLIANCE FOR  
PERMANENT ACCESS

# Overview of Innovation Project SCIDIP-ES and APA Tools for ESA LTDP Domain

David Giaretta APA

APARSEN-EGI Community Workshop on Managing,  
Computing and Preserving Big Data for Research

4-6 March 2014

Amsterdam



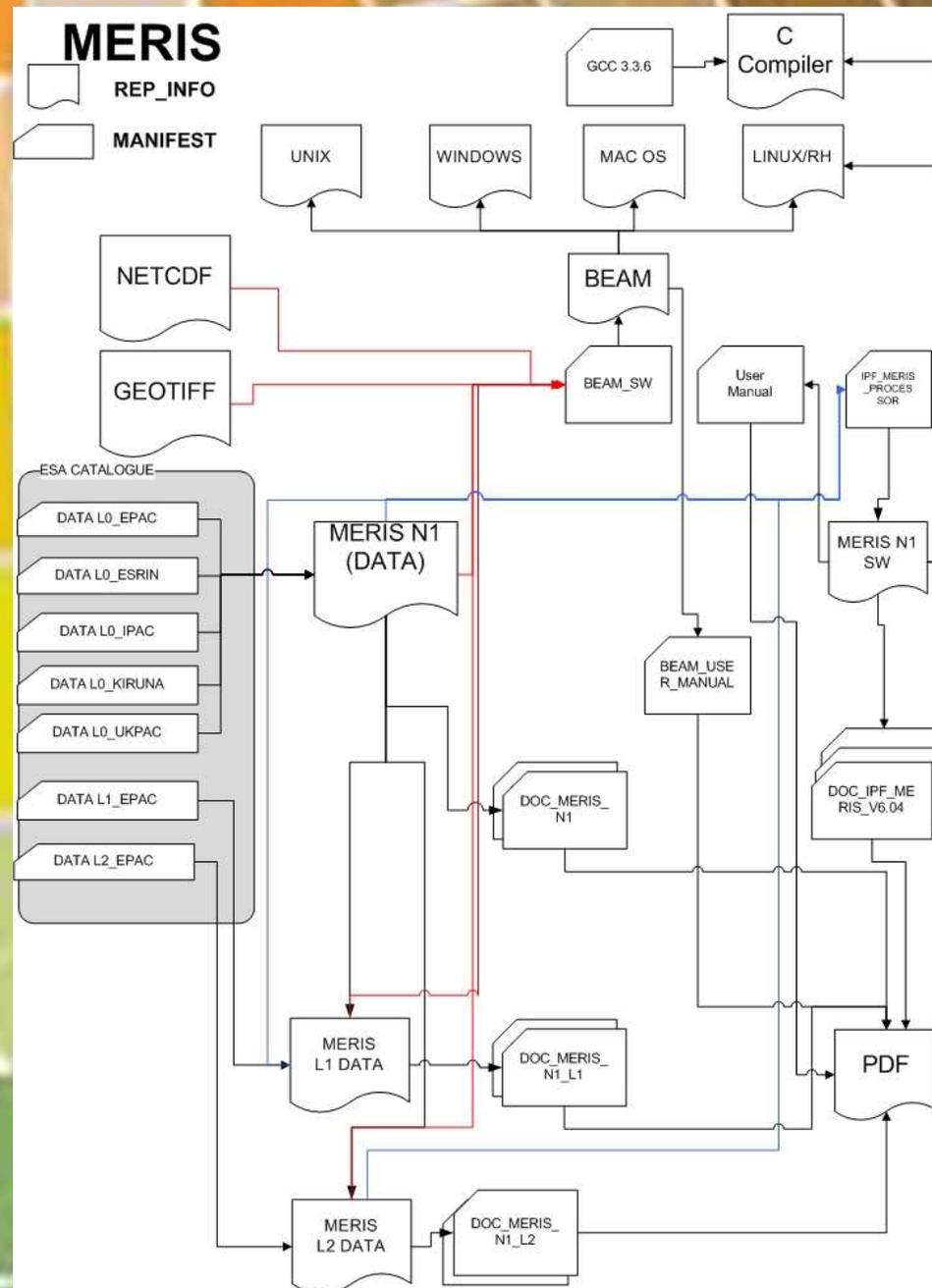
This work is made available under the terms of the Creative Commons Attribution-ShareAlike 3.0 license, <http://creativecommons.org/licenses/by-sa/3.0/>.



# Preserving digitally encoded information

- Ensure that digitally encoded information are **understandable** and **usable** over the long term
  - Long term could start at just a few years
  - Chain of preservation
- Need to do something because things become “unfamiliar” over time
- But the same techniques enable use of data which is “unfamiliar” right now

# MERIS example



# When things changes

- We need to:

- Know something has changed
- Identify the implications of that change
- Decide on the best course of action for preservation
- What RepInfo we need to fill the gaps
  - Created by someone else or creating a new one
- If transformed: how to maintain data authenticity
- Alternatively: hand it over to another repository
- Make sure data continues to be usable

Orchestration Service

Gap Identification Service

Preservation Strategy Tk

RepInfo Registry Service

Authenticity Toolkit

Storage Service

RepInfo Toolkit

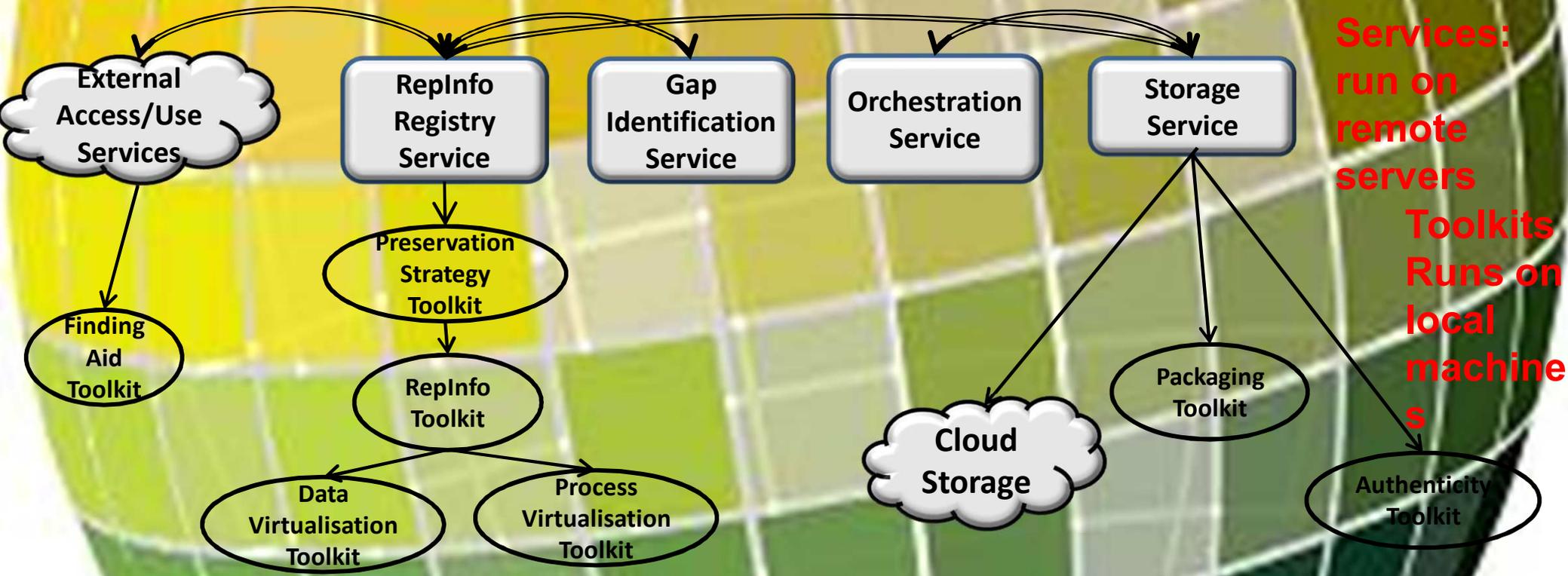
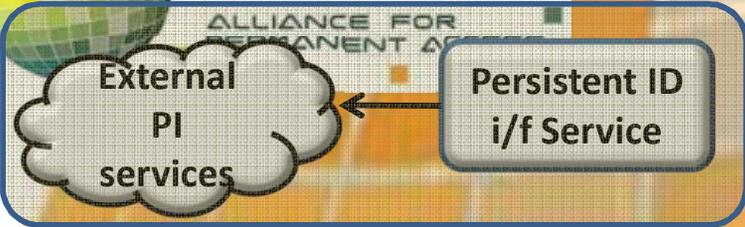
Data Virtualisation Toolkit

Process Virtualisation Toolkit

# SCIDIP-ES



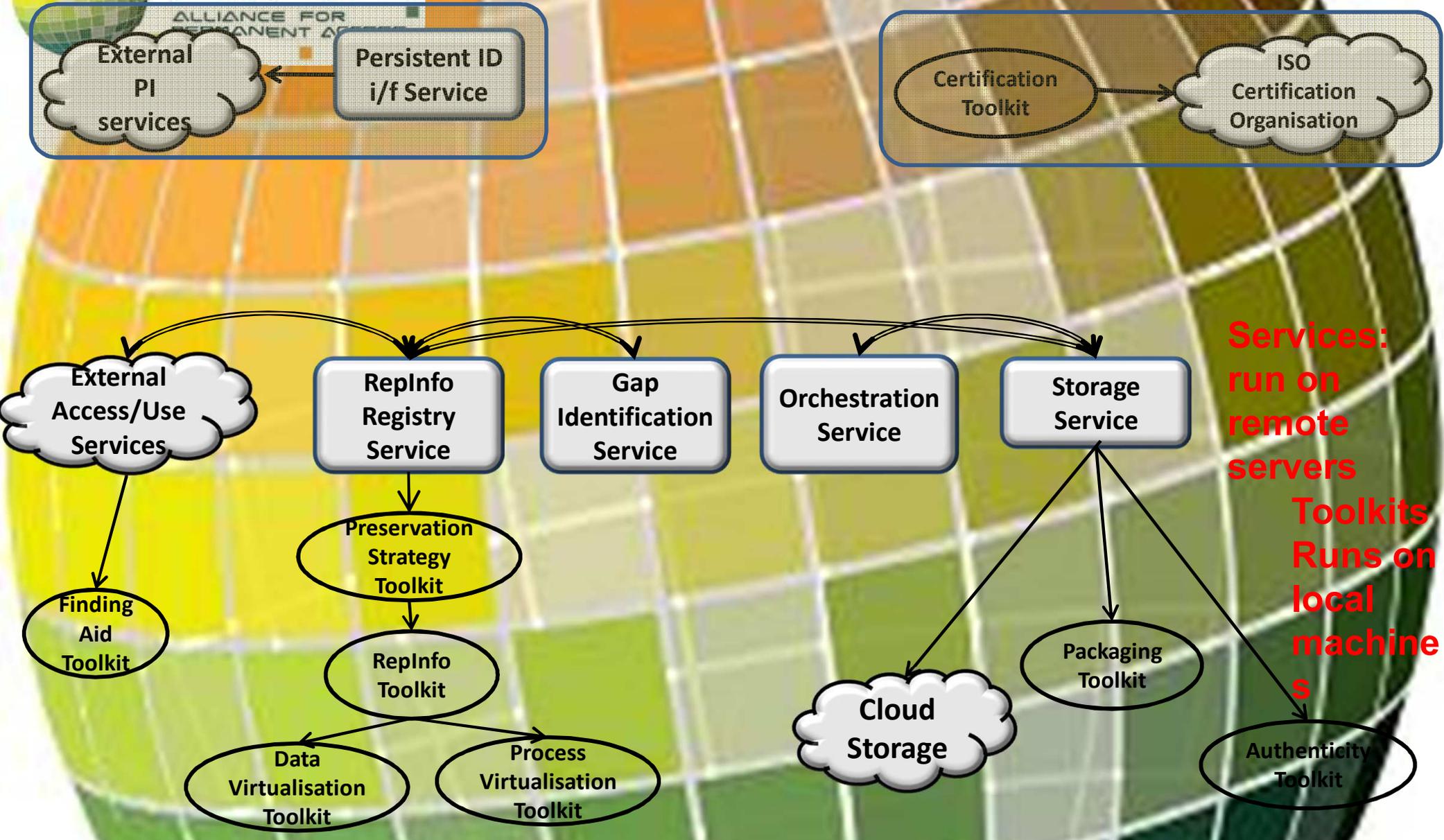
ALLIANCE FOR  
PERMANENT ACCESS



- Open Source software
- These SUPPLEMENT what repositories do (can be customised for repositories)
- Make it easier for repositories to do preservation – share the effort

Threat	Requirement for solution
Users may be unable to understand or use the data e.g. the semantics, format, processes or algorithms involved	<p><b>RepInfo toolkit, Packager and Registry</b> – to create and store Representation Information.</p> <p>In addition the <b>Orchestration Manager</b> and <b>Knowledge Gap Manager</b> help to ensure that the RepInfo is adequate.</p>
Non-maintainability of essential hardware, software or support environment may make the information inaccessible	<p><b>Registry</b> and <b>Orchestration Manager</b> to exchange information about the obsolescence of hardware and software, amongst other changes.</p> <p>The Representation Information will include such things as software source code and emulators.</p>
The chain of evidence may be lost and there may be lack of certainty of provenance or authenticity	<p><b>Authenticity toolkit</b> will allow one to capture evidence from many sources which may be used to judge Authenticity.</p>
Access and use restrictions may make it difficult to reuse data, or alternatively may not be respected in future	<p><b>Packaging toolkit to package</b> access rights policy into AIP</p>
Loss of ability to identify the location of data	<p><b>Persistent Identifier</b> system: such a system will allow objects to be located over time.</p>
The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future	<p><b>Orchestration Manager</b> will, amongst other things, allow the exchange of information about datasets which need to be passed from one curator to another.</p>
The ones we trust to look after the digital holdings may let us down	<p><b>Certification toolkit</b> to help repository manager capture evidence for ISO 16363 Audit and Certification</p>

# SCIDIP-ES



**Services:**  
run on  
remote  
servers

**Toolkits:**  
Runs on  
local  
machines

- These SUPPLEMENT what repositories do (customised for repositories)
- Make it easier for repositories to do preservation – share the effort



**APA**  
ALLIANCE FOR  
PERMANENT ACCESS

**END**

# When things changes

- We need to:

- Know something has changed
- Identify the implications of that change
- Decide on the best course of action for preservation
- What RepInfo we need to fill the gaps
  - Created by someone else or creating a new one
- If transformed: how to maintain data authenticity
- Alternatively: hand it over to another repository
- Make sure data continues to be usable

Orchestration Service

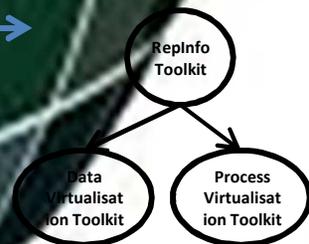
Gap Identification Service

Preservation Strategy Tk

RepInfo Registry Service

Authenticity Toolkit

Storage Service





# Problems when preserving data

- Preserve?
- Preserve what?
- For how long?
- How to test?
- Which people?
- Which organisations?
- How well?
- Metadata? – **What kind? How much?**

# Basic preservation activities

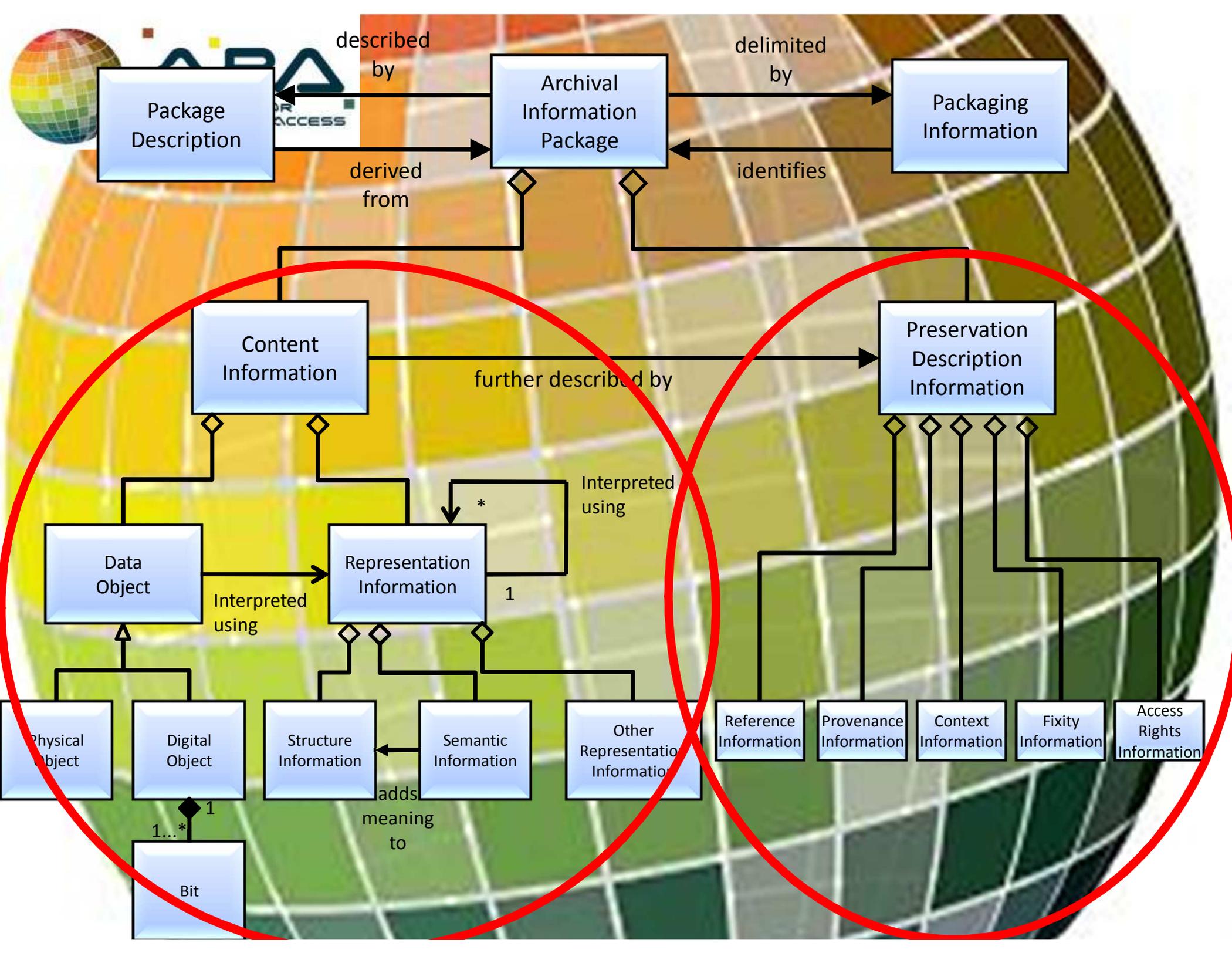
Libraries say:

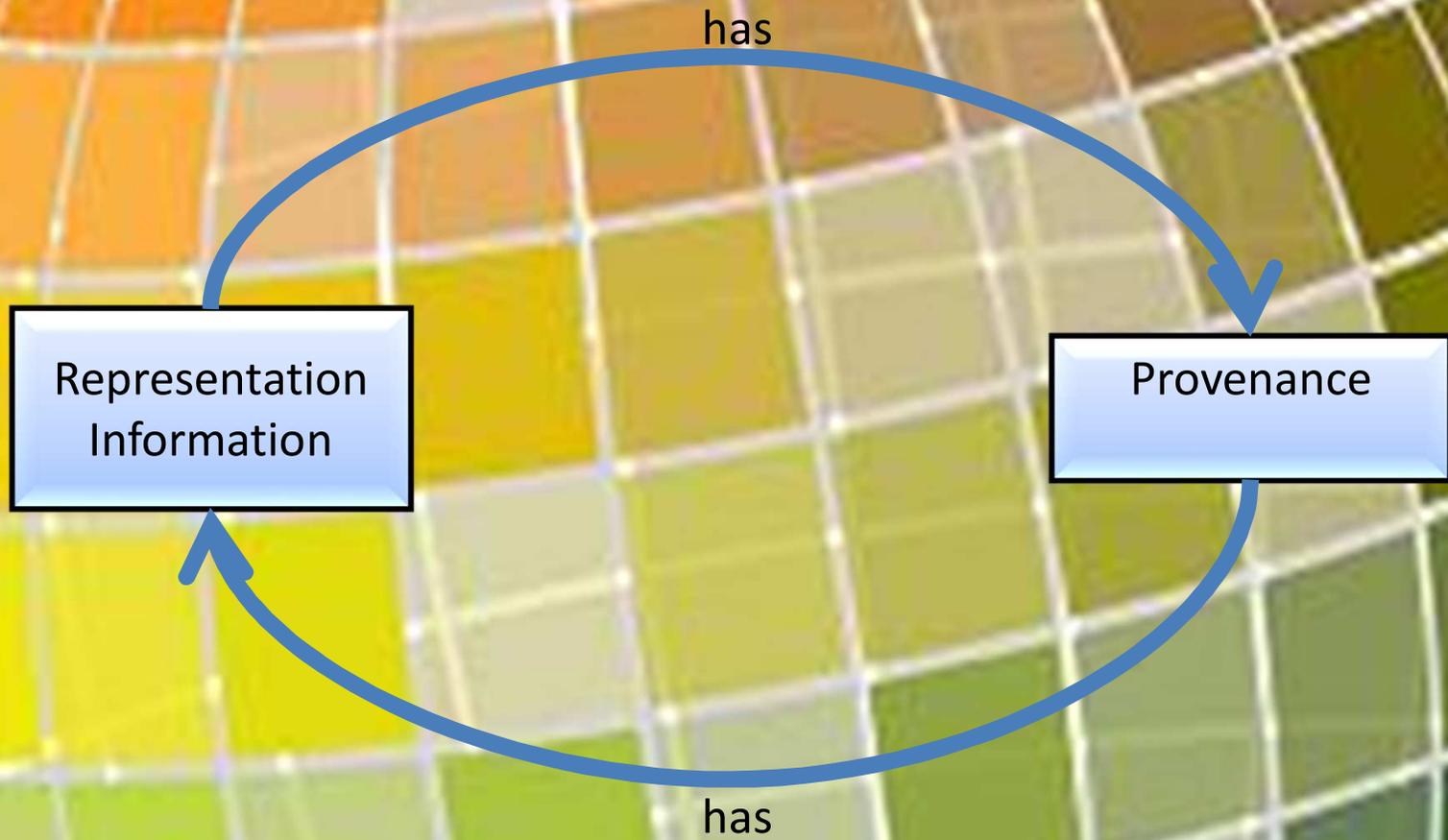
- “Emulate or migrate”

- Can repeat what has been done before
- BUT
- Cannot use new applications

- Convert to format which new software can use
- BUT
- What if there are many software systems?

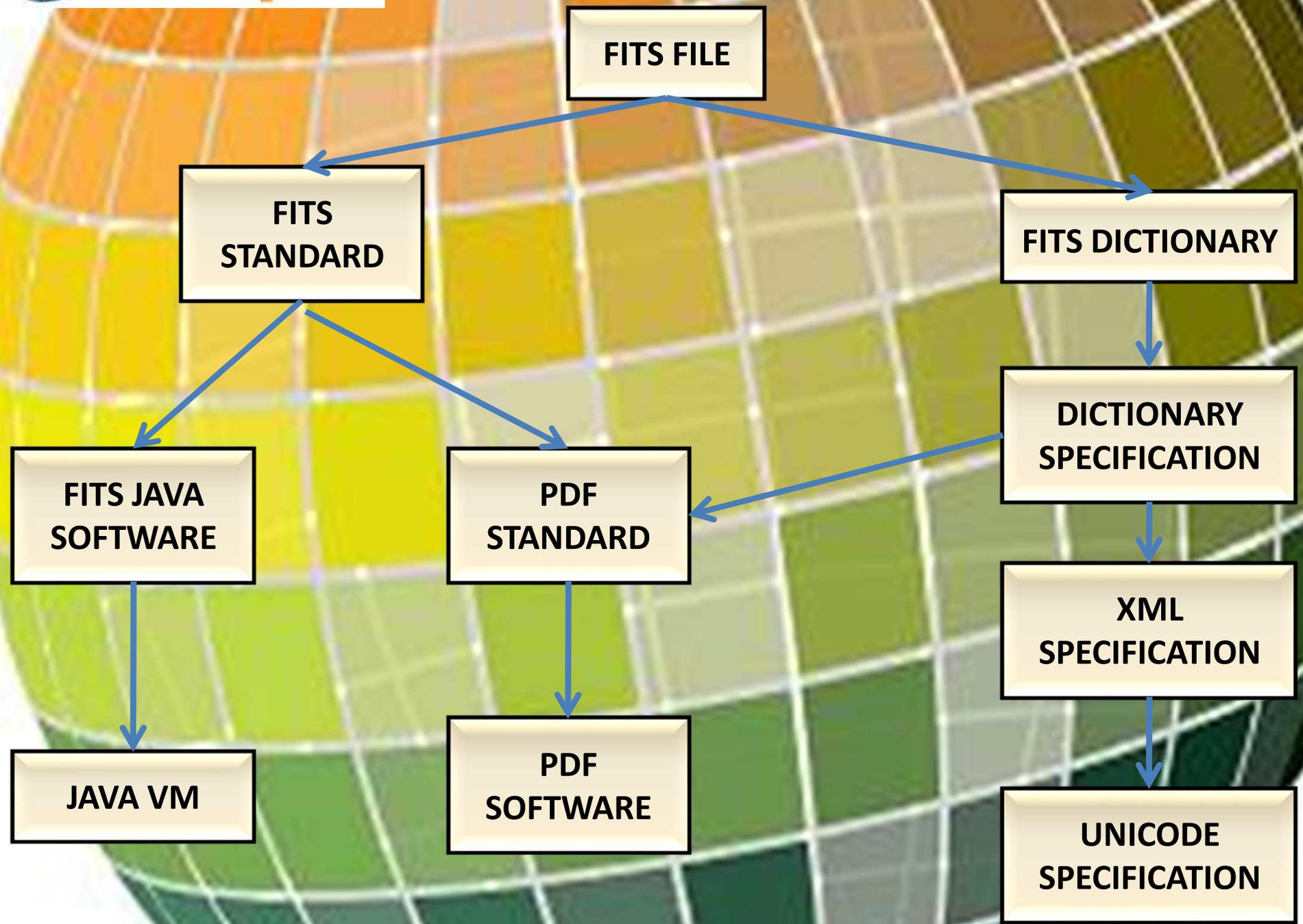
- Works well with data only in special cases
  - Can repeat what was done before instead of new things
- Does not help with building cross-disciplinary communities







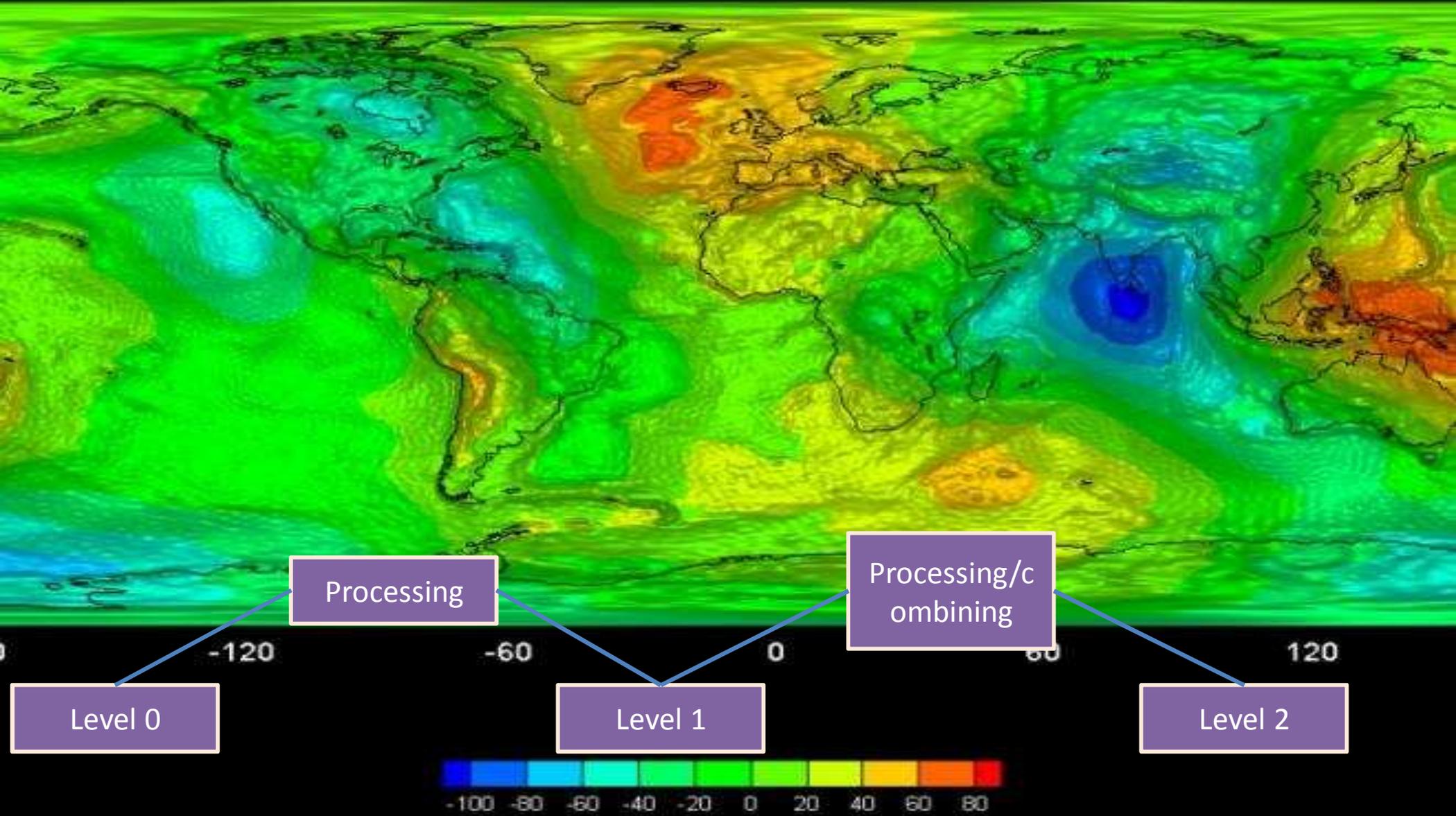
# Rep Info Network



Contains numbers – need  
meaning

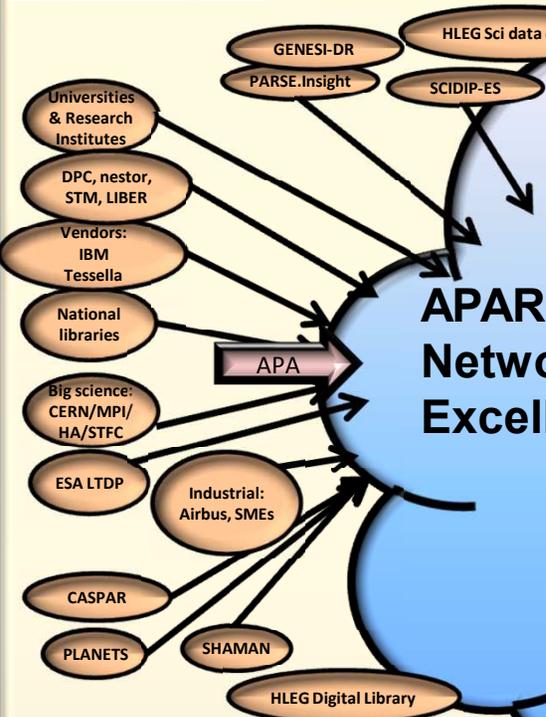
latitude	longitude	Ozone	Time
132	50	34.9	12/03/1999
178	50	45	12/03/1999
190	50	78	12/03/1999

...to be combined and processed to get this



# APARSEN

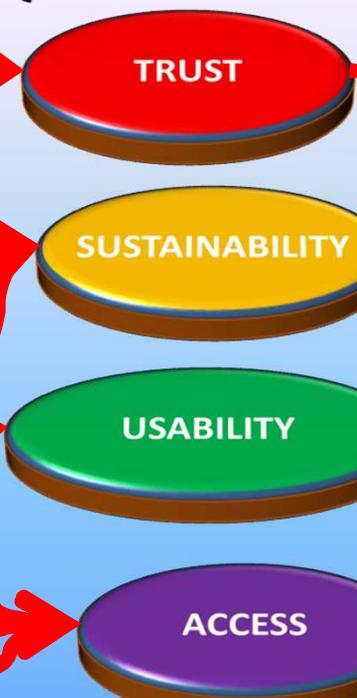
E-Infrastructures (Interoperability and preservation)



**APARSEN  
Network of  
Excellence**



**RESEARCH SILOS**



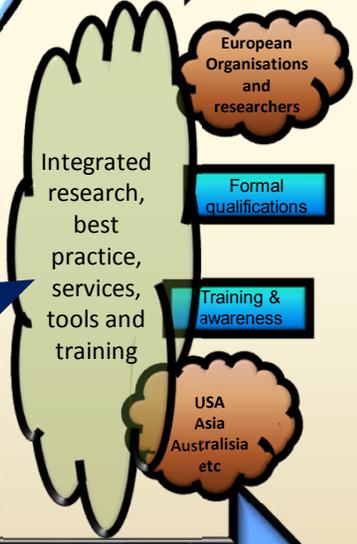
**INTEGRATED TOPICS**



**Common vision allows coherent approaches to training and practice**

**Research provides tools and techniques which feed into e-infrastructure**

**Virtual Centre of Excellence**



**APARSEN in brief**

- Create common vision for digital preservation research
- de-fragment existing efforts in academia, commerce, research labs, libraries etc

Research (Digital Preservation)

