DP Knowhow: Open Archival Information Systems (OAIS) in ISO 14721

Matthias Hemmje (FTK), David Giaretta (APA)

APARSEN-EGI-Community-Forum Training on Data Preservation

22nd of May 2014

Helsinki, Finland

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/.

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

Problems when preserving data

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

Difficulties in digital preservation

- Many different terminologies
- Many different views of preservation
- Many different kinds of digital objects
  - Documents
  - Data
  - …… and new types of objects
- Tools and Services
  - Which ones work for which digital objects?
  - Which tools/techniques fit together?
  - How to integrate new tools
- Consistent training needed
- Risks vs Cost
- Who can you trust?

OAIS – ISO 14721

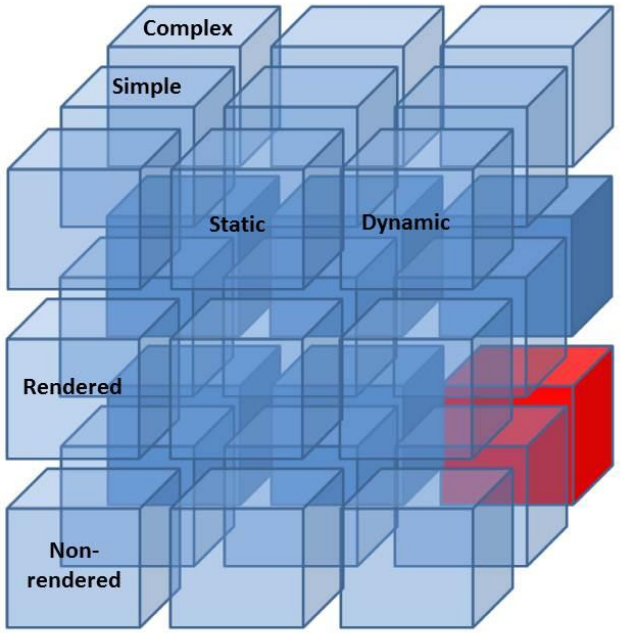
Need a consistent, coherent approach to digital preservation - APARSEN.

Need an Audit and Certification system – ISO 16363

Preservation techniques

For each technique

- look for evidence – what evidence?
- must at least make sure we consider different types of data
  - rendered vs non-rendered
  - composite vs simple
  - dynamic vs static
  - active vs passive
- must look at all types of th



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

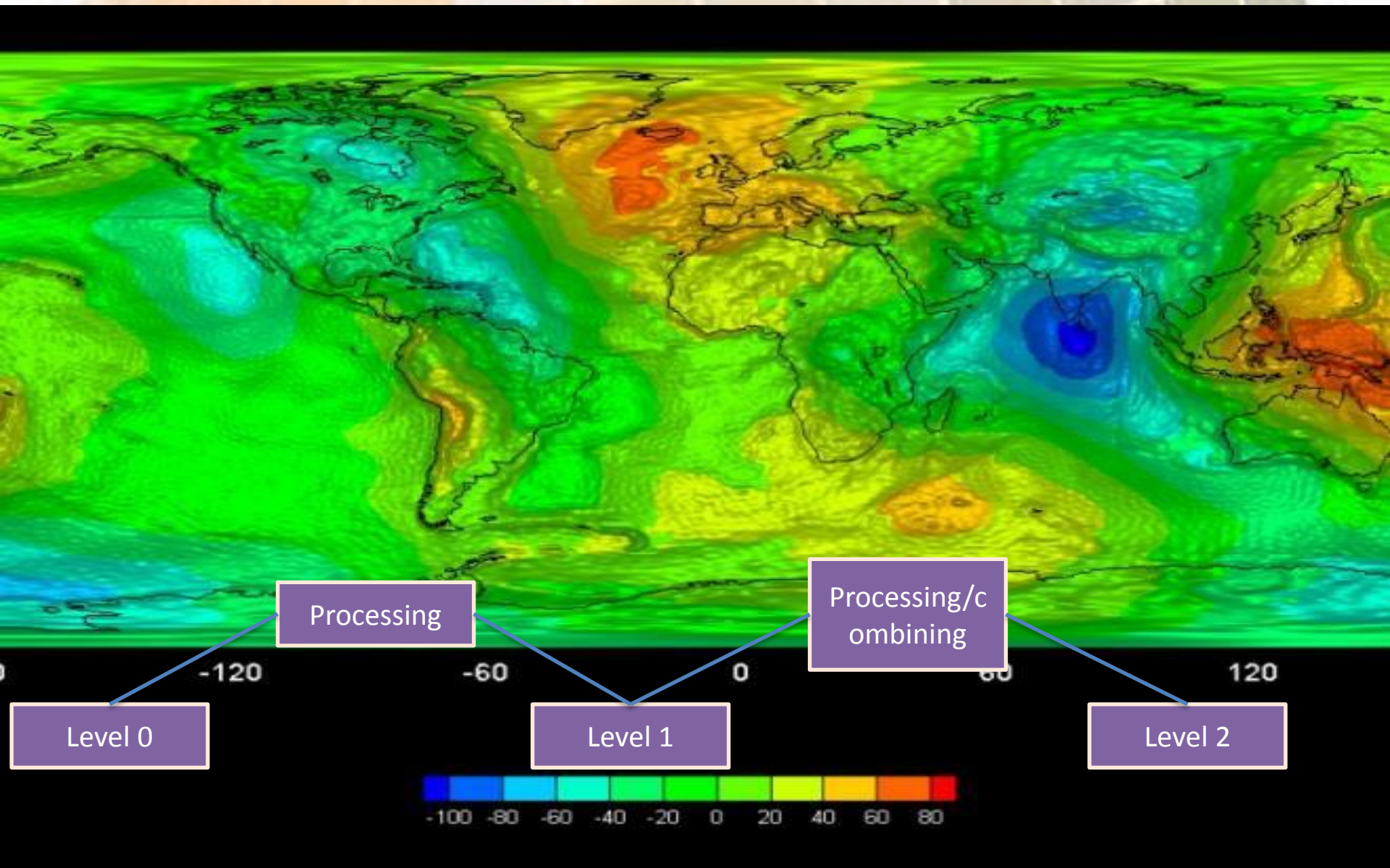
- What if there are many software systems?

  - 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

Contains numbers – need meaning

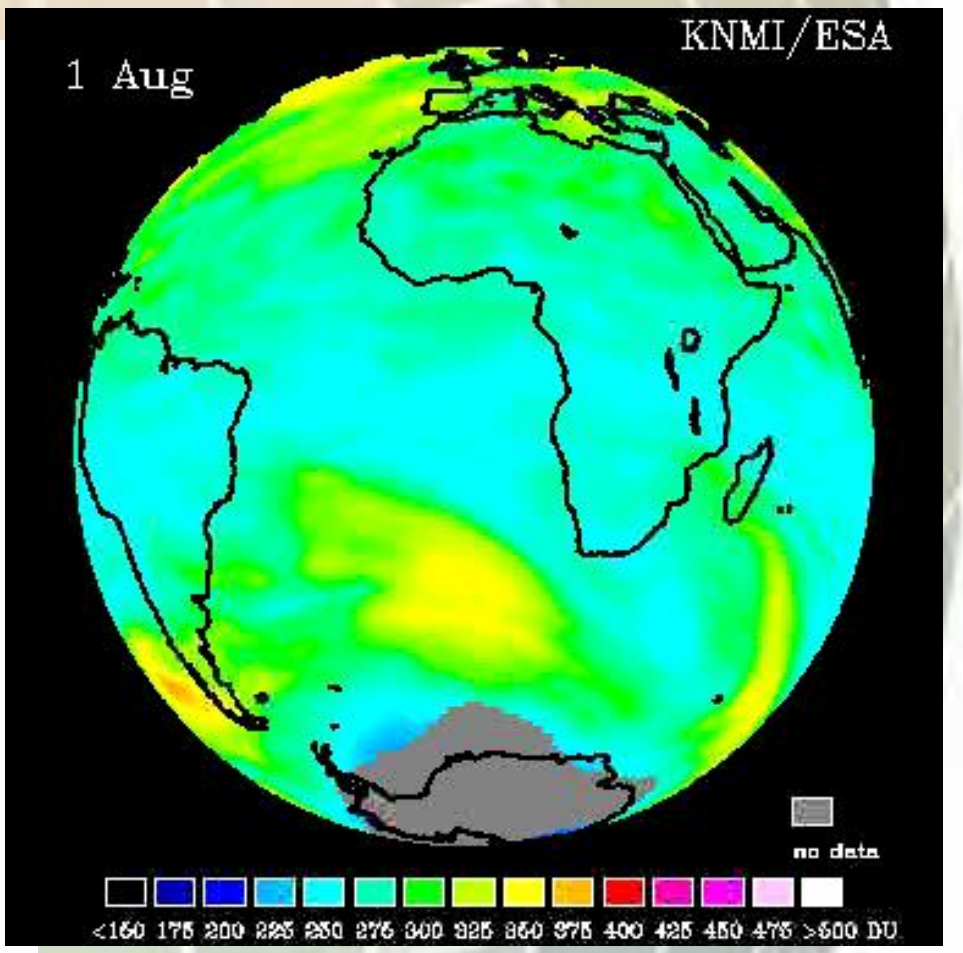
latitude	longitude	Ozone	Time
132	50	34.9	12/03/1999
178	50	45	12/03/1999
190	50	78	12/03/1999

...to be combined and processed to get this



...or this

GOME-NOPREGO ozone profiles, date 2001090100
10^12 [molec cm-3] iso-surface
SP


KNMI/ESA
1 Aug
no data
<160 175 200 225 250 275 300 325 350 375 400 425 450 475 >500 DU

Additional technique: add Representation Information

- Descriptions of the digitally encoded object
- Ideal description allows a machine to extract information

OAIS Information Model: Representation Information

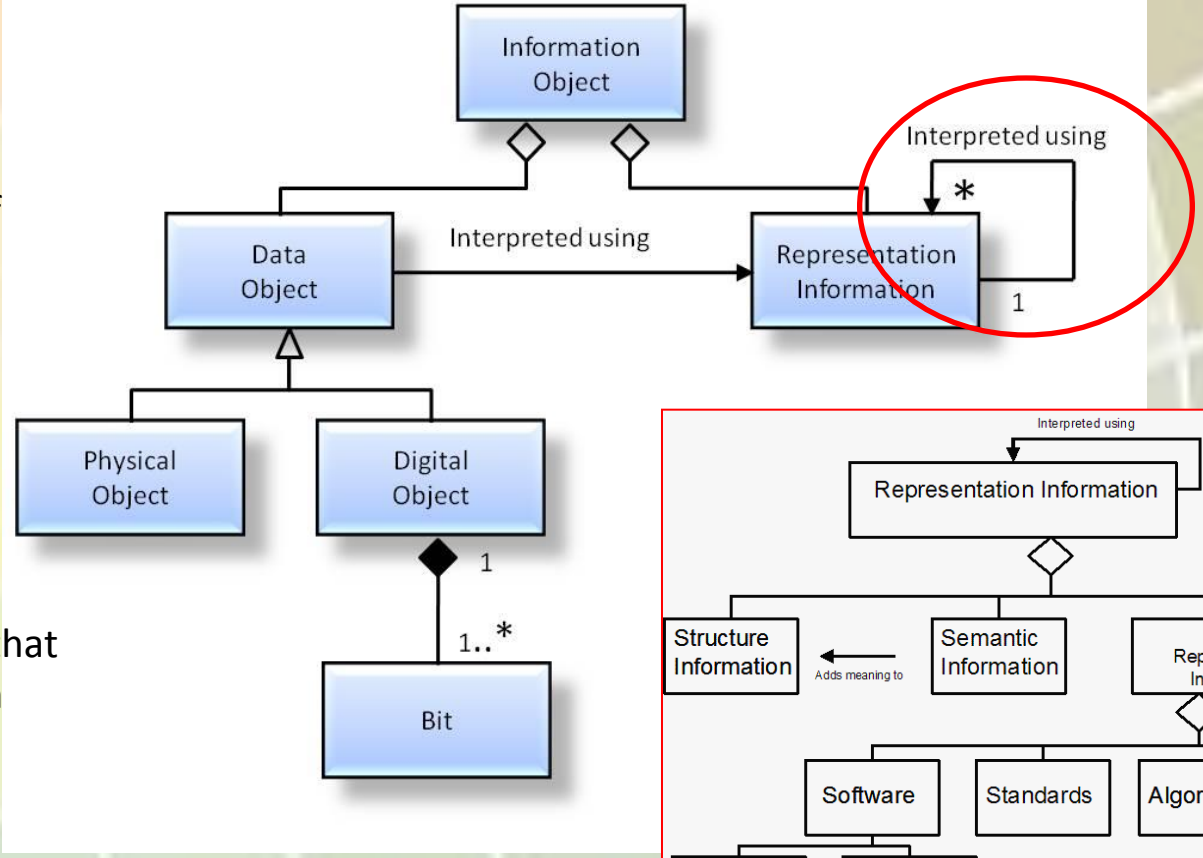
The Information Model is key

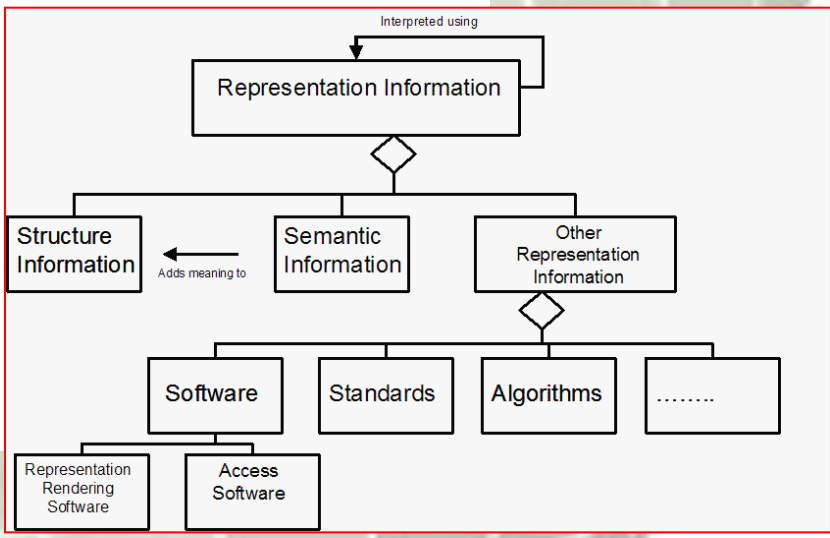
Recursion ends at KNOWLEDGEBASE of the DESIGNATED COMMUNITY

(this knowledge will change over time and region)

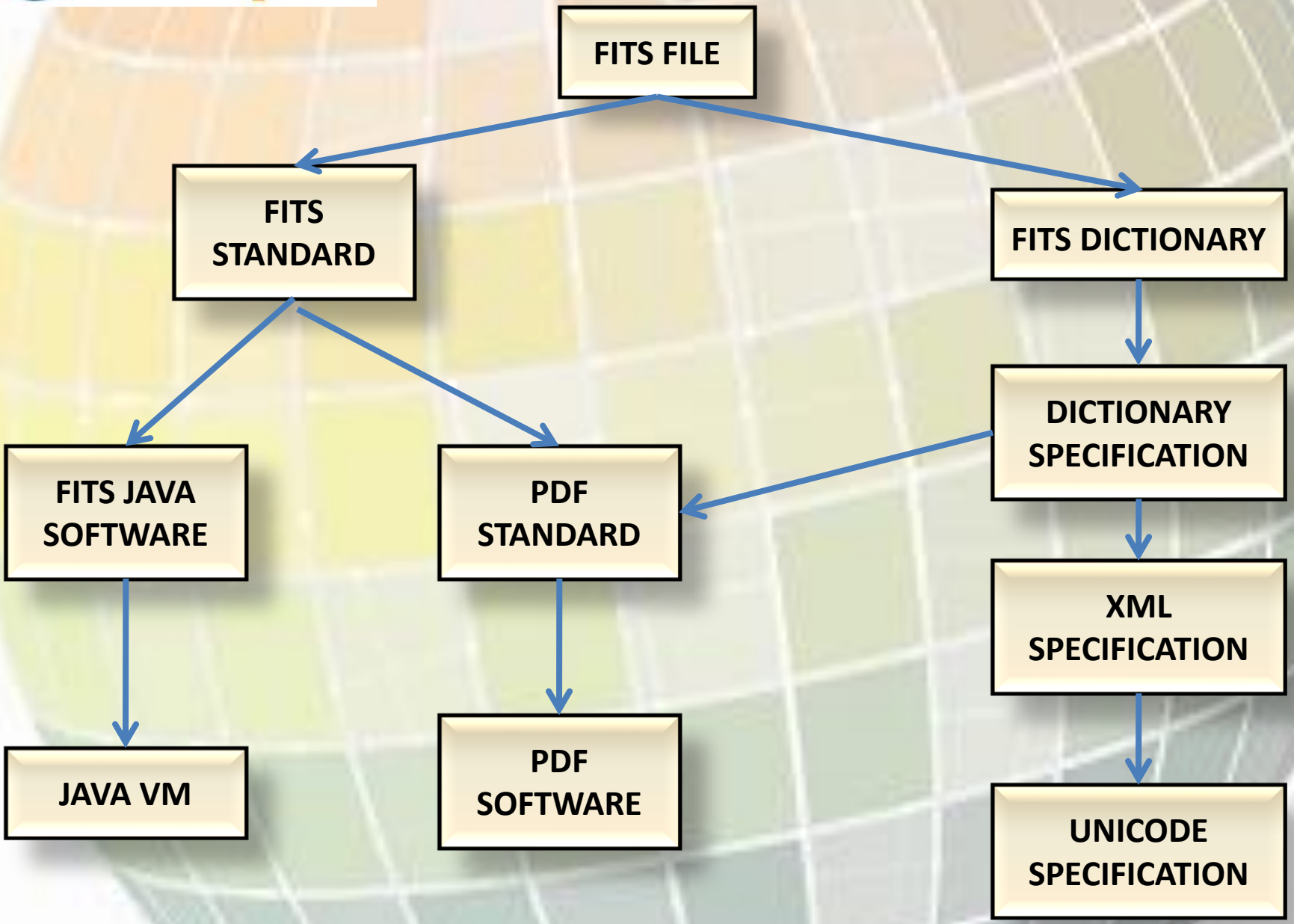
Does not demand that ALL Representation Information be collected at once.

A process which can be tested





Rep Info Network



AND – be prepared to Hand-over

- Preservation requires funding
- Funding for a dataset (or a repository) may stop
- Need to be ready to hand over everything needed for preservation
  - OAIS (ISO 14721) defines “Archival Information Package (AIP).
  - Issues:
    - Storage naming conventions
    - Representation Information
    - Provenance
    - ….