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

Significant Properties, Authenticity, Provenance, Representation Information and OAIS

David Giaretta

iPRES, San Francisco

5&6 October 2009

Information Society and Media

Authors

- **David Giaretta[1], Brian Matthews[1], Juan Bicarregui[1], Simon Lambert[1], Mariella Guercio[2], Giovanni Michetti[2] and Donald Sawyer[3]**
-
- [1]Science & Technology Facilities Council, Rutherford Appleton Lab., Didcot, Oxon, UK
- [2]University of Urbino, Italy
- [3]VIE, Inc., USA

Outline

- Introduction
- OAIS
- Discussion of definitions of Significant Properties
- New OAIS definitions
- Use of Significant Properties
- Summary

Introduction

- The term “Significant Properties” has been given a variety of definitions and used in various ways over the past several years.
- The relationship between Significant Properties and the OAIS term Representation Information has been a puzzle.
- Examples of Significant Properties of data do not seem to be available

Aim

- To describe the way in which the update to OAIS introduces concepts to clarify the situation and
  - provides a clear definition and name for Significant Properties
  - links these to Authenticity and Provenance
  - shows the relationship to Representation Information
  - provides examples covering all types of digital objects – data as well as documents

We claim this is consistent with the actual use of the Significant Properties and does not invalidate the previous pieces of work but rather provides a clear and consistent view of the concept.

Outline

- Introduction
- **OAIS**
- Discussion of definitions of Significant Properties
- New OAIS definitions
- Use of Significant Properties
- Summary

OAIS

- Version 1 did not use the term Significant Property
  - Not surprising since it came out in 2002
- Talks about **many** concepts including
  - Representation Information
  - Authenticity (not treated well)
  - Various preservation strategies
    - More than just migrate or emulate

New version of OAIS

- Currently undergoing review – see http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Overview.aspx
- Revised many definitions in the light of suggestions and discussions

Outline

- Introduction
- OAIS
- **Discussion of definitions of Significant Properties**
- New OAIS definitions
- Use of Significant Properties
- Summary

Definitions (1)

The CEDARS project [7] defined Significant Properties as

- *"those characteristics [both technical, intellectual, and aesthetic] agreed by the archive or by the collection manager to be the most important features to preserve over time".*

Definitions (2)

Sergeant [13] on the other hand proposed that

- *"Significant Properties are those attributes of an object that constitute the complete (for the intended Consumer) intellectual content of that object"*

- However the example given of
  - *Significant Properties for an e-thesis of*
  - *the complete text, including divisions into chapters and sections*
  - *the layout and style - particular fonts and spacing are essential*
  - *Diagrams*
  - *(perhaps web adverts are not Significant for our e-journals).*

Definitions (3)

The OCLC/RLG Working Group on Preservation Metadata [12] proposed the definition:

- *"Properties of the Content Data Object's rendered content which must be preserved or maintained during successive cycles of the preservation process"*

Definitions (4)

Hedstrom and Lee [6] defined Significant Properties as

- *"those properties of digital objects that affect their quality, usability, rendering, and behaviour".*

In that paper they have a number of links to the OAIS Reference Model, for example

- *"whether or not colour, for example, is a significant property of the given digital object or collection will depend on the extent to which colour features affect the quality and usability of the preserved object for a designated community",*

- and
  - *"decisions about which Significant Properties to maintain will depend on institutional priorities, anticipated use, knowledge of the designated community, the types of materials involved, and the financial and technical resources available to the repository"*

Definitions (5)

Within the InSPECT project, Wilson [4] defines Significant Properties in a similar fashion as

- *"the characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability and meaning of the objects".*

Problems

- New, different, definitions come out quite frequently!
- No example for scientific data
- Comparing the various definitions
  - only Wilson, 2007 includes “meaning” in its definition,
  - Sergeant, 2002 includes what might be interpreted as a more ambitious phrase “complete (for the intended Consumer) intellectual content”;
  - Wilson, 2007 is the only one to include “accessibility”.
  - Both Hedstrom and Lee, 2002, and Wilson, 2007 include “usability” in their definitions which is plausible but hard to see, for example, with “redness”.
  - The terms “appearance” and “experienced” is used in Rothenberg and Bikson, 1999 while Hedstrom and Lee, 2002 includes “rendering” and “behaviour”;
  - OCLC/RLG Working Group on Preservation Metadata, 2002 refers to “rendered content” and, as noted above,
  - the example in Sergeant, 2002 makes it fairly clear that the rendering is the main concern.

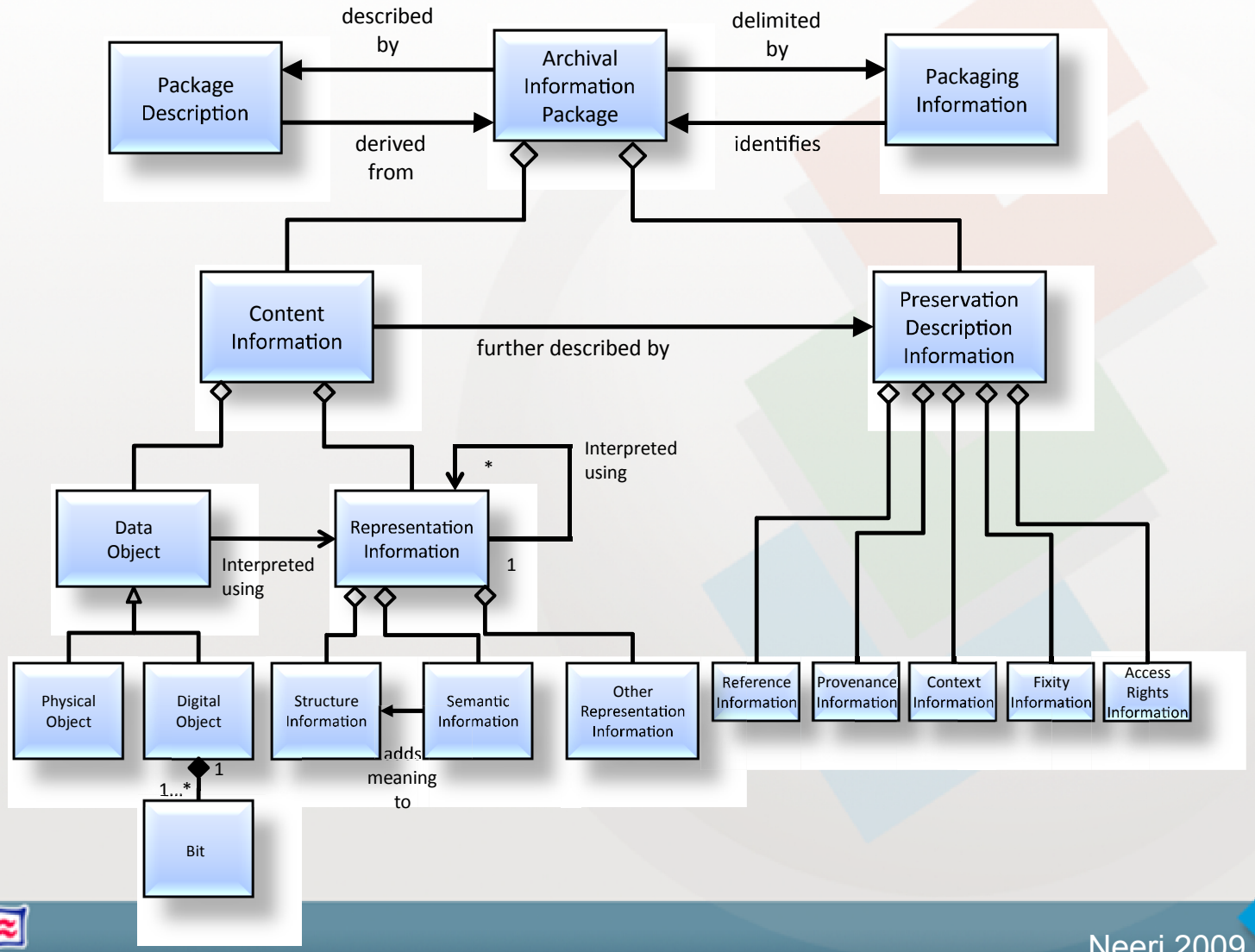
Role of Significant Properties

- Important for Migrations (actually Transformations) of digital object:
  - Current usage focuses on RENDERING but:
    - What about digital objects where rendering is not a concern - DATA?
    - How do we know which Transformations are acceptable?
    - How do we know that the new object is”authentic”?

Outline

- Introduction
- OAIS
- Discussion of definitions of Significant Properties
- **New OAIS definitions**
- Use of Significant Properties
- Summary

OAIS Archival Information Package (AIP)



Authenticity

– *"the degree to which a person (or system) may regard an object as what it is purported to be. The degree of Authenticity is judged on the basis of evidence"*

Provenance

- *"The information that documents the history of the Content Information. This information tells the origin or source of the Content Information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. The archive is responsible for creating and preserving Provenance Information from the point of Ingest, however earlier Provenance Information should be provided by the Producer. Provenance Information adds to the evidence to support Authenticity."*

Types of Migration

Operations which do not change the bit sequences

- **Refreshment**: A Digital Migration where a media instance, holding one or more AIPs or parts of AIPs, is replaced by a media instance of the same type by copying the bits on the medium used to hold AIPs and to manage and access the medium. As a result, the existing Archival Storage mapping infrastructure, without alteration, is able to continue to locate and access the AIP.
- **Replication**: A Digital Migration where there is no change to the Packaging Information, the Content Information and the PDI. The bits used to convey these information objects are preserved in the transfer to the same or new media-type instance. Refreshment is also a Replication, but Replication may require changes to the Archival Storage mapping infrastructure.

Operations which change the bit sequences

- **Repackaging**: A Digital Migration where there is some change in the bits of the Packaging Information.
- **Transformation**: A Digital Migration where there is some change in the Content Information or PDI bits while attempting to preserve the full information content.

Types of Transformation

- A **Reversible Transformation** occurs when the new representation defines a set (or a subset) of resulting entities that are equivalent to the resulting entities defined by the original representation. This means that there is a one-to-one mapping back to the original representation and its set of base entities
- A **Non-Reversible Transformation** occurs when a Reversible Transformation cannot be guaranteed.

OAIS view:

- **New version defines a number of concepts**
  - **However NOT *Significant Property* – because that would add yet another definition to the confusion**
- **Transformational Information Property**: An Information Property whose preservation is regarded as being necessary but not sufficient to verify that the Non-Reversible Transformation has adequately preserved information content. This could be important as contributing to evidence about Authenticity. Such Information Properties will need to be associated with specific Representation Information, including Semantic Information, to denote how they are encoded and what they mean. (Note that the term 'significant property', which has various definitions in the literature, is sometimes used in a way that is consistent with it being a Transformational Information Property).

Role in Authenticity

- Someone (the curator) has to verify that the Transformation was acceptable
  - the curator presumably would have satisfied himself or herself that the new object resulting from the transformation continued to “maintain” authenticity.
    - This may have been done by, for example, checking that the words were correct between the Word file and the PDF file; that the rendering of the pages was reasonably consistent between the two versions; that text which had been emphasised in the Word version by highlighting or by changing colour, was also a emphasized in some appropriate way in the PDF version.

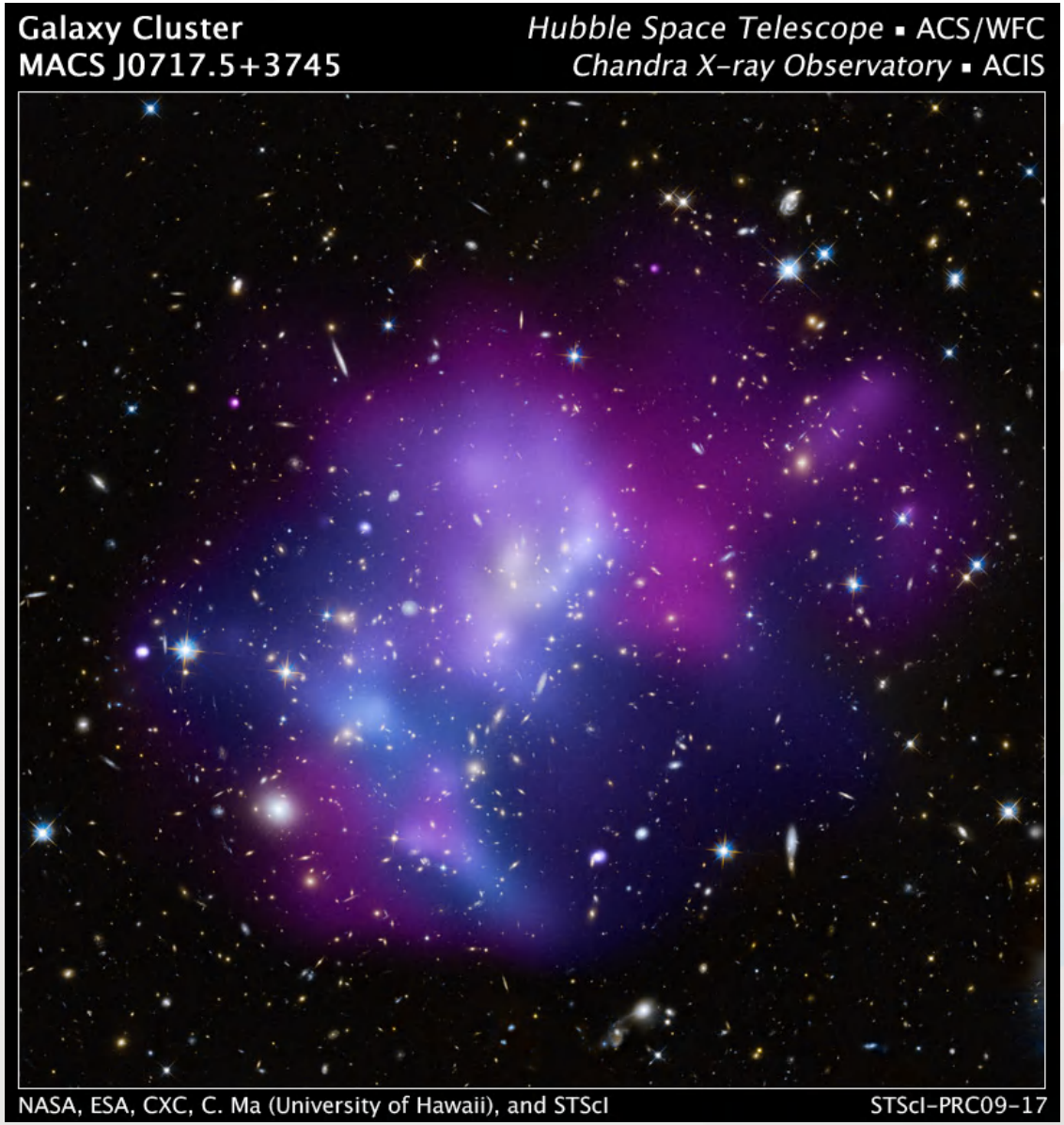
Consistent with current usage

Outline

- Introduction
- OAIS
- Discussion of definitions of Significant Properties
- New OAIS definitions
- **Use of Significant Properties**
- Summary

Example for Data (1)

- Example: FITS file converted to a CDF file.
  - The bit sequences will have been changed extensively.
  - How could a curator have satisfied himself or herself that the object as transformed had not lost required information content and therefore was still being adequately preserved and could be regarded as being as authentic as the FITS file.


Galaxy Cluster
MACS J0717.5+3745
Hubble Space Telescope ▪ ACS/WFC
Chandra X-ray Observatory ▪ ACIS
NASA, ESA, CXC, C. Ma (University of Hawaii), and STScI
STScI-PRC09-17

Example for Data (continued)

- The FITS file might contain an image;
  - The CDF file should contain a similar image.
  - However just comparing the two images rendered on screens would be inadequate for scientific purposes.
  - Instead the curator would need to be satisfied, for example,
    - that the data values of the pixel elements were identical in the two images at corresponding points;
    - that the co-ordinates associated with each pixel in the two images were identical, for example the same latitude and longitude;
    - that the units associated with the numerical values were the same in both images.
- Note that the curator would not have to understand the what the data means to do these tests
  - Not necessarily linked to Designated Community

Outline

- Introduction
- OAIS
- Discussion of definitions of Significant Properties
- New OAIS definitions
- Use of Significant Properties
- **Summary**

Summary

- Definitions of Significant Property are many and various – rather confusing
- OAIS
  - introduces ***Transformational Information Property***
  - Shows that these are important for
    - The chain of Authenticity
    - The choice of Transformation
- Hope that the new term and its (soon to be) ISO definition is widely used in future

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

References

1. CCSDS. (2002), Reference model for an Open Archival Information System (OAIS). Retrieved from: http://public.ccsds.org/publications/archive/650x0b1.pdf
2. OAIS update (at the time of writing under CCSDS review), http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf
3. Knight, G., 2008, Framework for the definition of significant properties. Retrieved from http://www.significantproperties.org.uk/documents/wp33-propertiesreport-v1.pdf
4. Wilson, A., 2007, Significant Properties Report. Retrieved from http://www.significantproperties.org.uk/documents/wp22_significant_properties.pdf
5. J. Rothenberg and T. Bikson, 1999, 'Carrying Authentic, Understandable and Usable Digital Records Through Time' report to the Dutch National Archives and Ministry of the Interior. Retrieved fromt http://www.digitaleduurzaamheid.nl/bibliotheek/docs/final-report_4.pdf
6. M. Hedstrom and C.A. Lee, “Significant properties of digital objects: definitions, applications, implications”, Proceedings of the DLM-Forum 2002. Retrieved from http://ec.europa.eu/transparency/archival_policy/dlm_forum/doc/dlm-proceed2002.pdf
7. Cedars project, http://www.leeds.ac.uk/cedars/
8. Investigating the Significant Properties of Electronic Content over time (InSPECT) http://www.significantproperties.org.uk/
9. The InterPARES project, http://www.interpares.org/
10. Wison, A., 2008, Significant Properties of Digital Objects, presented at “What to preserve? Significant Properties of Digital Objects”. Retrieved from http://www.dpconline.org/docs/events/080407sigpropsWilson.pdf
11. DELOS Digital Preservation Testbed. Retrieved from http://www.ifs.tuwien.ac.at/dp/testbed.html
12. OCLC/RLG Working Group on Preservation Metadata, 2002, Preservation Metadata and the OAIS Information Model, A Metadata Framework to Support the Preservation of Digital Objects. Retrieved from http://www.oclc.org/research/projects/pmwg/pm_framework.pdf
13. Derek Sergeant, 2002, Interpretation of the OAIS Model. Retrieved from http://www.erpanet.org/events/2002/copenhagen/presentations/dmserpanet.ppt
14. CASPAR Access Model, http://www.casparpreserves.eu/Members/cclrc/Deliverables/report-on-oais-access-model/at_download/file especially section 2.
15. Michael Factor, Ealan Henis, Dalit Naor, Simona Rabinovici-Cohen, Petra Reshef, Shahar Ronen, IBM Research Lab in Haifa, Israel and Giovanni Michetti, Maria Guercio, University of Urbino, Authenticity and Provenance in Long Term Digital Preservation: Modelling and Implementation in Preservation Aware Storage, TaPP '09. First Workshop on the Theory and Practice of Provenance. San Francisco, 23 February 2009, http://www.usenix.org/event/tapp09/tech/full_papers/factor/factor.pdf
16. CASPAR Conceptual Model, http://www.casparpreserves.eu/Members/cclrc/Deliverables/caspar-conceptual-model-phase-1-1/at_download/file
17. Giaretta, D., 2007, The CASPAR Approach to Digital Preservation, The International Journal of Digital Curation, Issue 1, Volume 2, http://www.ijdc.net/index.php/ijdc/article/viewFile/29/18
18. CASPAR – Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval. See http://www.casparpreserves.eu
19. Mike Coyne, David Duce, Bob Hopgood, George Mallen, Mike Stapleton. The Significant Properties of Vector Images. JISC report, 27 November 2007. http://www.jisc.ac.uk/media/documents/programmes/preservation/vector_images.pdf
20. Mike Coyne, Mike Stapleton. The Significant Properties of Moving Images. JISC report, 26 March 2008. http://www.jisc.ac.uk/media/documents/programmes/preservation/spmovimages_report.pdf
21. Brian Matthews, Brian McIlwrath, David Giaretta, Esther Conway. The Significant Properties of Software: A Study. JISC report, March 2008 http://www.jisc.ac.uk/media/documents/programmes/preservation/spsoftware_report_redacted.pdf
22. Kevin Ashley, Richard Davis, Ed Pinsent. Significant Properties of E-learning Objects. JISC report, March 2008. http://www.jisc.ac.uk/media/documents/programmes/preservation/spelos_report.pdf
23. PARADIGM project, Workbook on Digital Private Papers. http://www.paradigm.ac.uk/workbook/preservation-strategies/file-properties.html

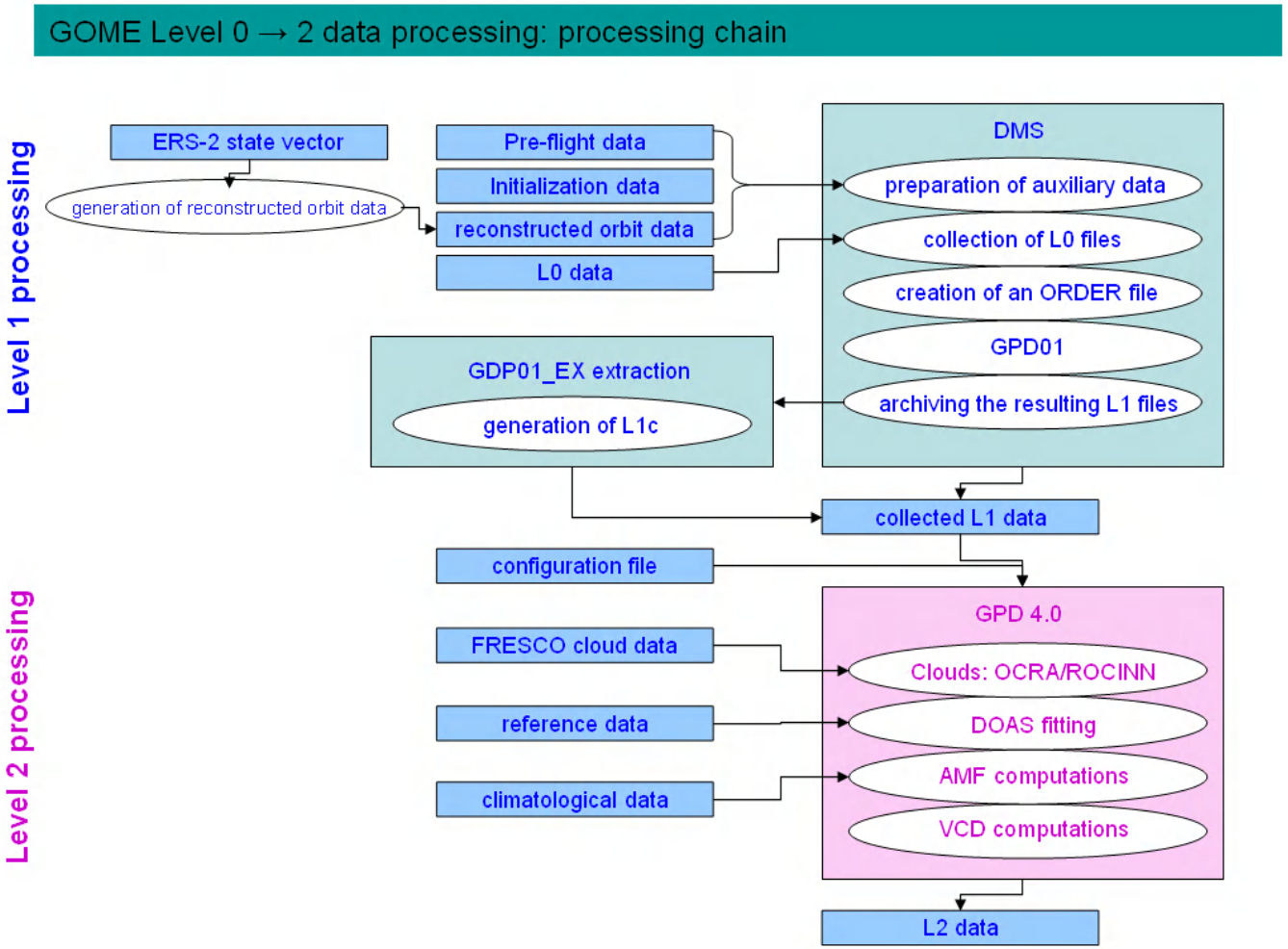
Information Society and Media

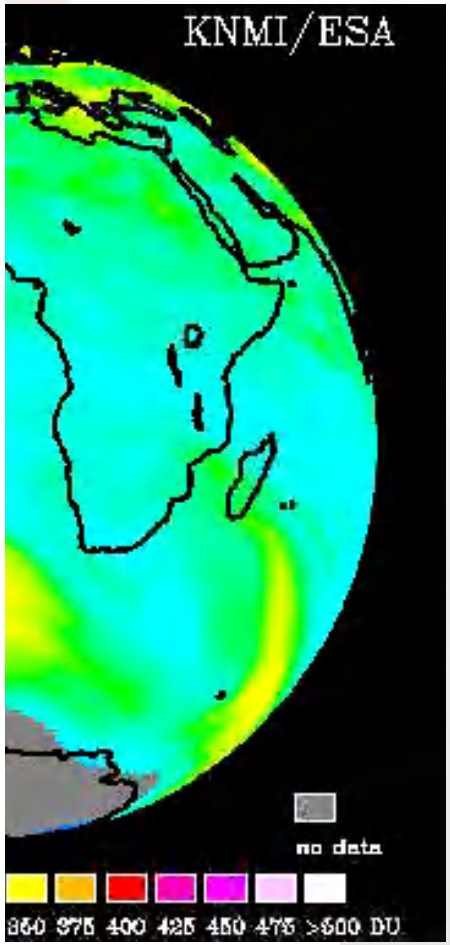
Links

- CASPAR:
  - ***www.casparpreserves.eu***
- Audit and certification:
  - ***wiki.digitalrepositoryauditandcertification.org***
- OAIS:
  - ***http://public.ccsds.org/publications/archive/650x0b1.pdf***
  - ***http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Overview.aspx***

END

Data…


GOME Level 0 → 2 data processing: processing chain
Level 1 processing
Level 2 processing
ERS-2 state vector
generation of reconstructed orbit data
Pre-flight data
Initialization data
reconstructed orbit data
L0 data
DMS
preparation of auxiliary data
collection of L0 files
creation of an ORDER file
GPD01
archiving the resulting L1 files
GDP01_EX extraction
generation of L1c
collected L1 data
configuration file
FRESCO cloud data
reference data
climatological data
GPD 4.0
Clouds: OCRA/ROCINN
DOAS fitting
AMF computations
VCD computations
L2 data


KNMI/ESA
no data
360 375 400 425 450 475 >600 DU