



D | C | C

Funded by:

JISC

Funded by:



Digital Object Semantics

Stephen Rankin
(on behalf of David Giaretta)

Overview

- Significant Properties
- OAIS Information Model
- Representation Information
- Significant Properties vs Representation Information
- Rendered Objects
- Semantics
- Example
- Conclusions



D | C | C



Significant Properties

- From the workshop description:
 - Significant properties are essential characteristics of a digital object which must be preserved over time for the digital object to remain accessible and meaningful.
 - Proper understanding of the significant properties of digital objects is critical to establish best practices and helps answer the fundamental question related to digital preservation: what to preserve?



D | C | C



Key OAIS (ISO 14721) Concepts

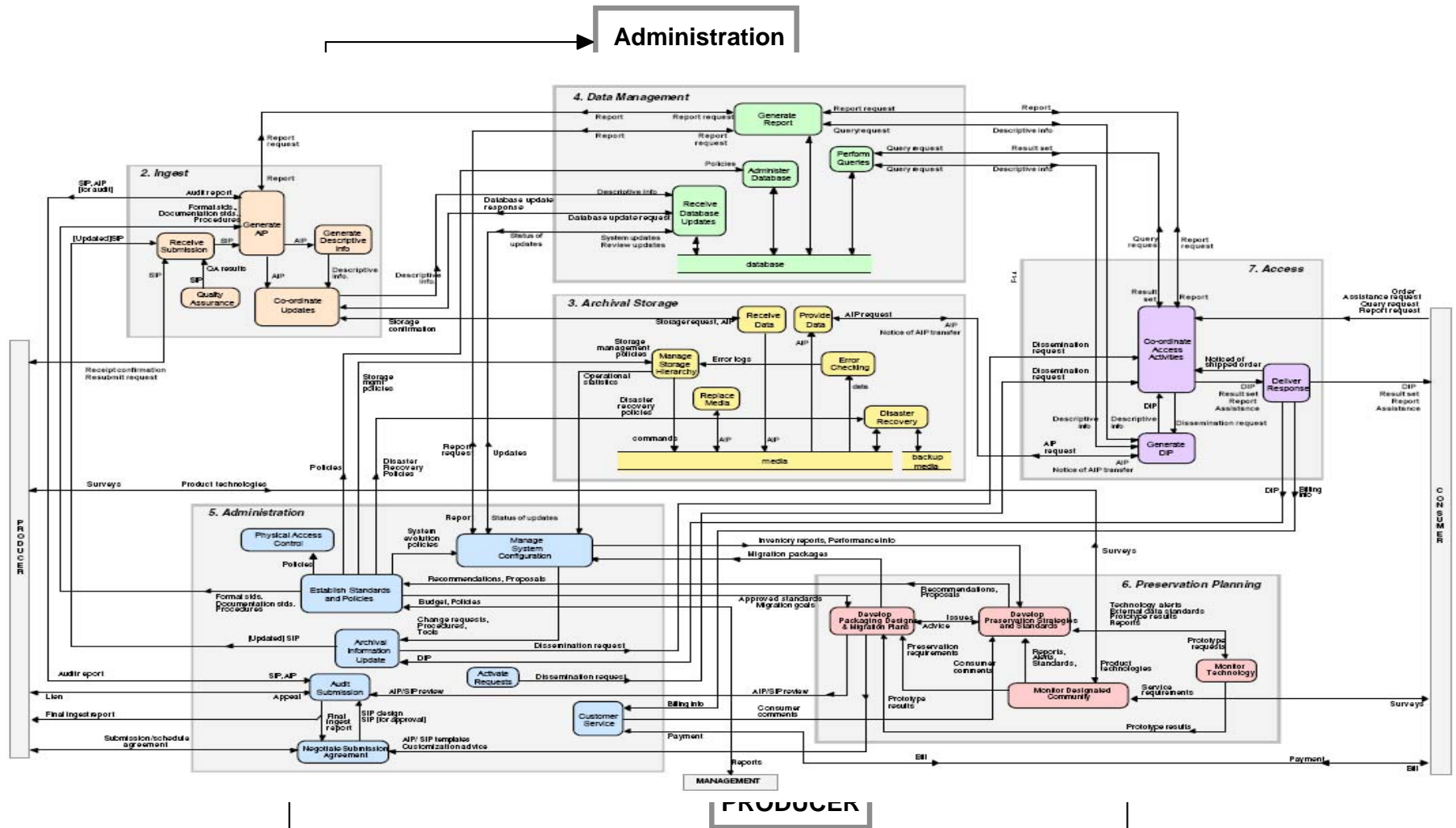
- Claiming “This is being preserved” is untestable
 - Essentially meaningless
- How can we make it testable?
 - Claim to be able to continue to “do something” with it
 - Understand/use
 - Need Representation Information
- Still meaningless...
 - Things are too interrelated
 - Representation Information potentially unlimited
 - Designated Community
- Plus many other concepts



| D | C | C



OAIS Functional Entities



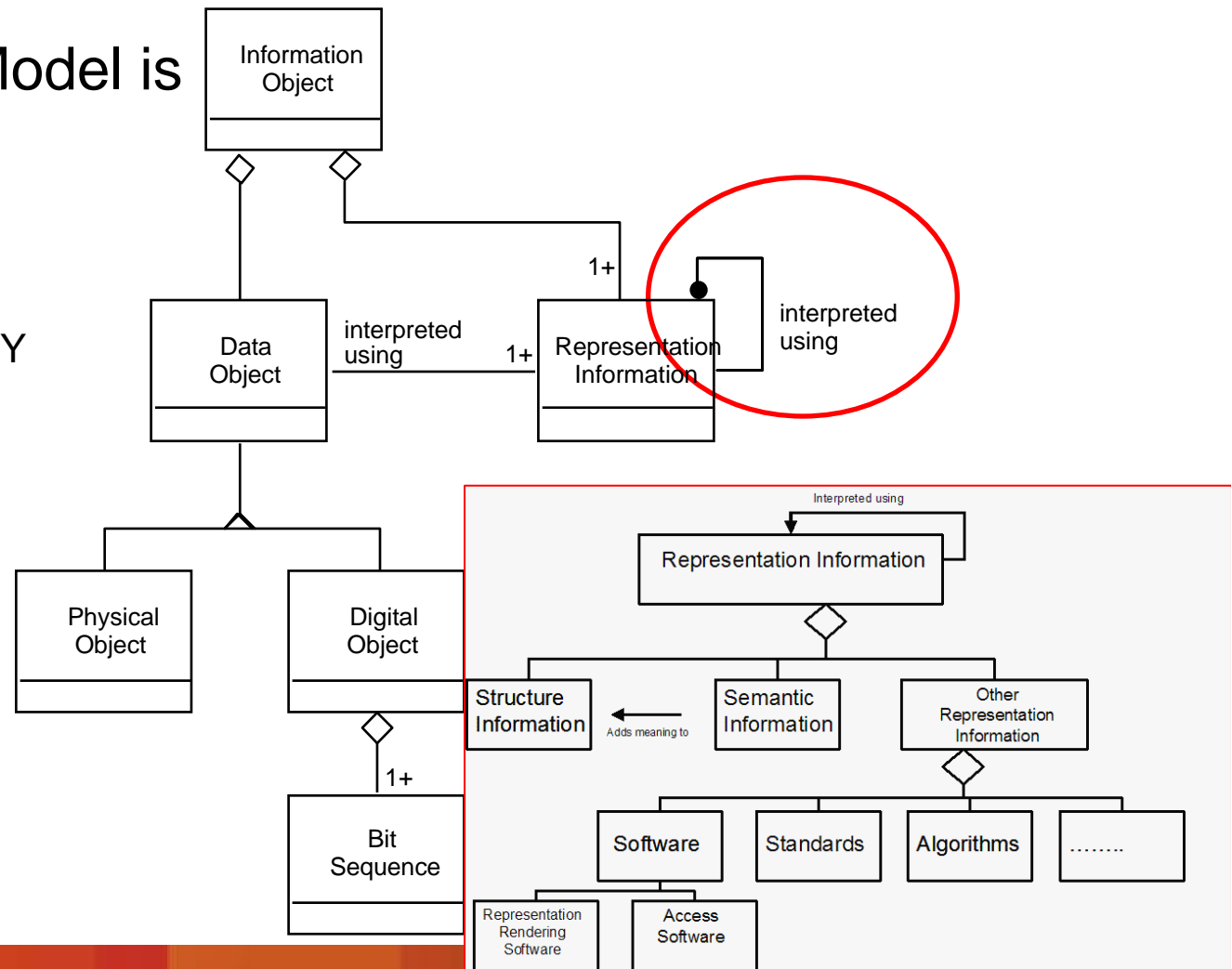
SIP = Submission Information Package
 AIP = Archival Information Package
 DIP = Dissemination Information Package



OAIS Information Model and Representation Information

The Information Model is key

Recursion ends at KNOWLEDGEBASE of the DESIGNATED COMMUNITY
(this knowledge will change over time and region)



Representation Information

- The information that maps a Data Object into more meaningful concepts.
- An example of Representation Information for a bit sequence which is a FITS file might consist of the FITS standard which defines the format plus a dictionary which defines the meaning of keywords in the file which are not part of the standard



D | C | C



Significant Properties and

Representation Information

- The term “Significant Properties” is used to indicate those properties of a Digital Object which needs to be preserved
- These often therefore will need to have specific Representation Information, usually either Structure or Other Representation Information, to denote how it is encoded.
- Examples
 - colour of text is GREEN
 - how is the colour encoded? 24 bits, colour system?
 - Software behaviour
 - E.g. what is needed to support emulation
 - Networks, performance usually neglected



D | C | C



Format “vs” Representation Information

- Format
 - IS Structural Representation Information
 - IS adequate for rendering a digital object
 - IS NOT adequate for understanding – especially data



| D | C | C



Just Format?

representation information rules

You have a file

JHOVE tells you it is WORD version 7

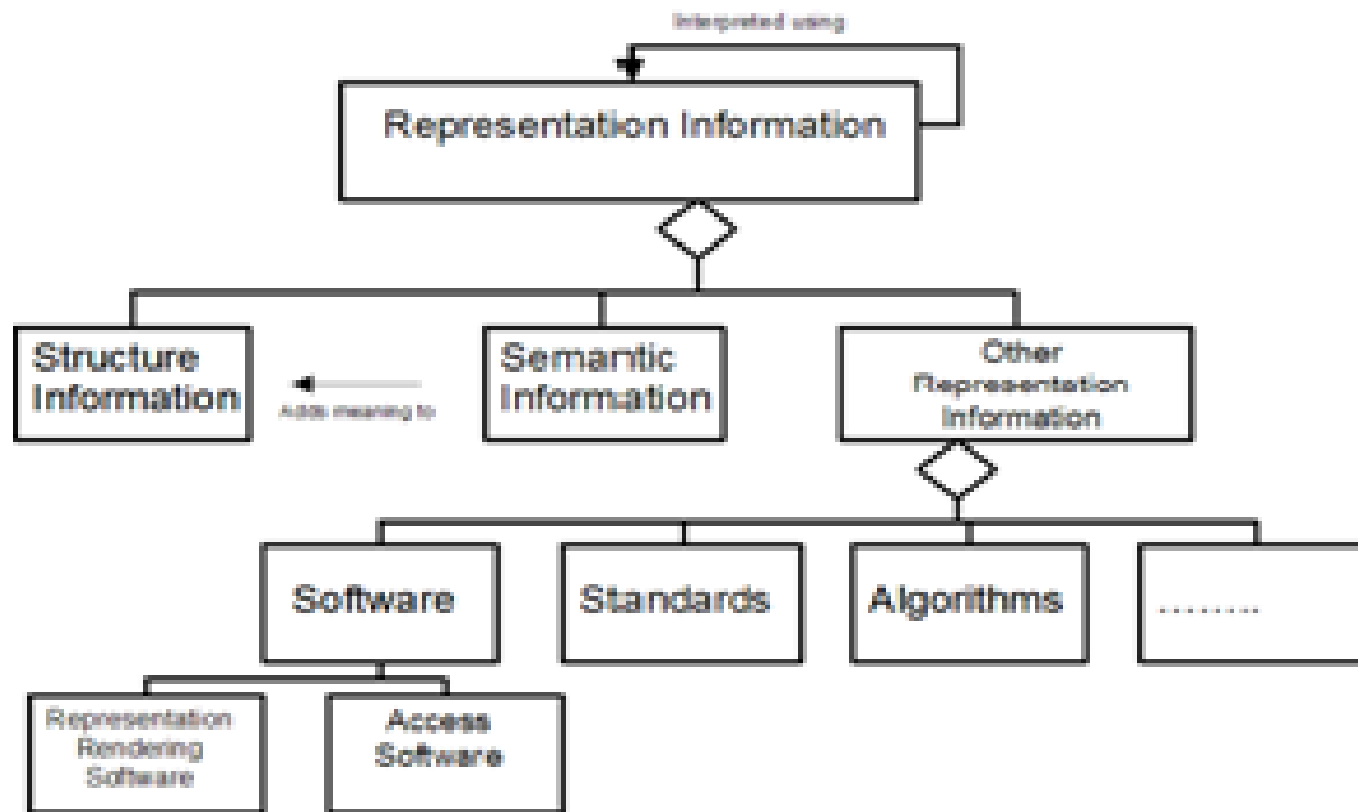
Format Registries – useful but not enough: formats can be used for multiple purposes e.g. audio files used to store configuration parameters



| D | C | C



OAIS Representation Information



D | C | C



Rendered Documents people

- Documents, articles, journals...
- Images
- Audio
- Video

A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing

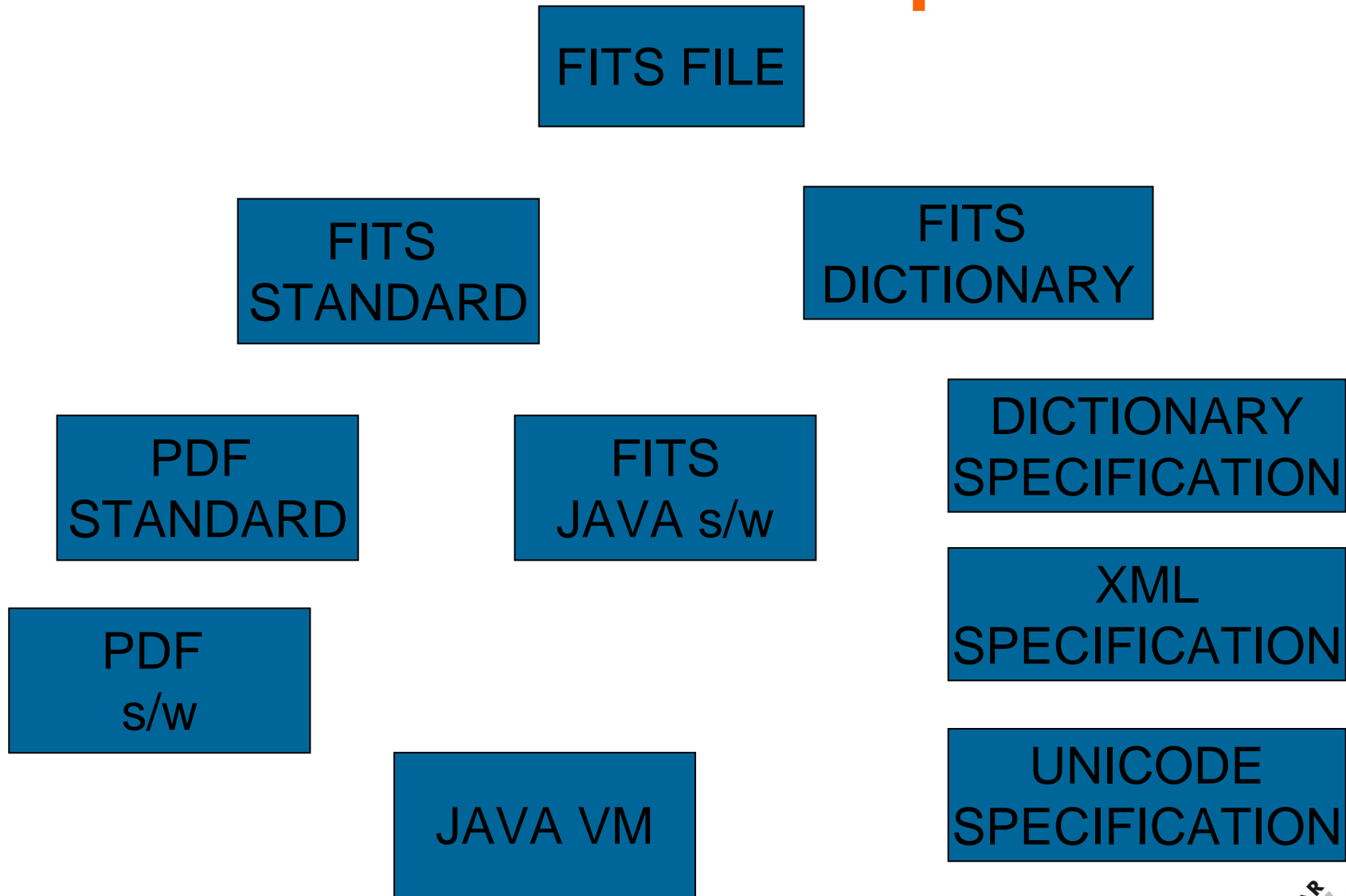
Semantics tends to be ignored



D | C | C



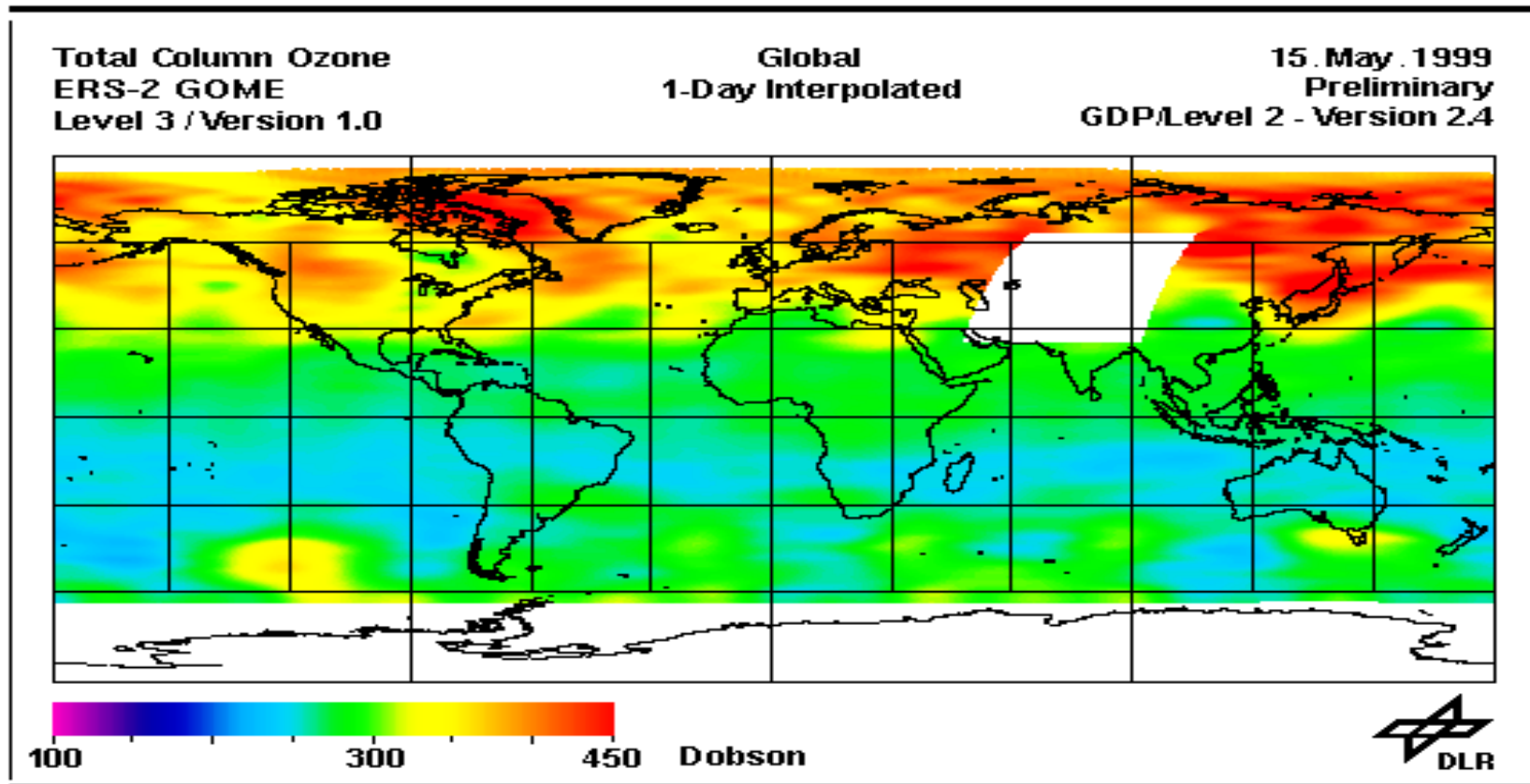
FITS file example



D | C | C



Data...



D | C | C



Effect of Transformations

- Transforming data as part of the process of preserving a piece of digitally encoded information:
 - Technical document in Word document → PDF
 - Scientific data in CDF → NetCDF
- The STRUCTURE Rep. Info. will change
- The SEMANTIC Rep. Info. will NOT change



| D | C | C



Conclusions

- Need to be able to preserve digitally encoded information
- Need Representation Information
 - Semantics as well as Structure and other things
- Significant Properties give some hint about SOME of the Representation Information which is needed
 - Usually Structure or Software Representation Information
- Much more Representation Information is needed
 - The amount of ReplInfo depends on the Designated Community – this demands clear definition
- Question: for whom, are the properties significant?



| D | C | C

