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
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?
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
OAIS Functional Entities

SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package
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.
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
Format “vs” Representation Information

• Format
  • IS Structural Representation Information
  • IS adequate for rendering a digital object
  • IS NOT adequate for understanding – especially data
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
XML enough?

<VOTABLE version="1.1"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://www.ivoa.net/xml/VOTable/v1.1 http://www.ivoa.net/xml/VOTable/v1.1"
 xmlns="http://www.ivoa.net/xml/VOTable/v1.1">
 <RESOURCE>
  <TABLE name="6dfgs_E7_subset" nrow="875">
   <PARAM arraysize="*" datatype="char" name="Original Source" value="http://www-wfau.roe.ac.uk/6dFGS/6dfgs_E7.fld.gz">
    <DESCRIPTION>URL of data file used to create this table.</DESCRIPTION>
   </PARAM>
   <PARAM arraysize="*" datatype="char" name="Comment" value="Cut down 6dfGS dataset for TOPCAT demo usage."/>
   <FIELD arraysize="15" datatype="char" name="TARGET">
    <DESCRIPTION>Target name</DESCRIPTION>
   </FIELD>
   <FIELD arraysize="11" datatype="char" name="DEC" unit="DMS">
    <DATA>
     <FITS>
      <STREAM encoding='base64'>
       U01NUEFxIICAgICAgICAgICAgICAgICAgICAgICBUIC8qU3RhbmRhcmQgRkIUUyBm
      </STREAM>
     </FITS>
    </DATA>
   </FIELD>
  </TABLE>
 </RESOURCE>
</VOTABLE>
OAIS Representation Information
A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing

Semantics tends to be ignored
FITS file example

- FITS FILE
- FITS STANDARD
- PDF STANDARD
- PDF s/w
- FITS JAVA s/w
- JAVA VM
- FITS DICTIONARY
- DICTIONARY SPECIFICATION
- XML SPECIFICATION
- UNICODE SPECIFICATION
Data...
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
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 RepInfo depends on the Designated Community – this demands clear definition
• Question: for whom, are the properties significant?