One Thing is Missing or Two Things are Confused
An Analysis of OAIS Representation Information

Simone Sacchi, Karen M. Wickett, Allen H. Renear, David Dubin {sacchi1, wickett2, renear, ddubin}@illinois.edu
Center for Informatics Research in Science and Scholarships — GSLIS — University of Illinois at Urbana-Champaign

We describe two alternative interpretations of OAIS Representation Information and show that both are flawed. The first is insufficient to formalize a model of preservation, and the second leads to category mistakes in conceptualizing the nature of digital artifacts.

**OAIS Information Model**
The OAIS Reference Model (CCSDS, 2002) claims that:

- **Data [Objects] interpreted using its Representation Information** yields Information

And provides the following definitions:

- **Data Object**: either a Physical Object or a Digital Object
- **Information**: any type of knowledge that can be exchanged
- **Representation Information**: information that maps a Data Object into more meaningful concepts

**The example**
The following example is offered as an illustration:

“The information stored within a CD-ROM file is expressed by the bits (the data) ... Assume the bits represent an ASCII table of numbers giving the coordinates of a location on the Earth measured in degrees latitude and East longitude. The Representation Information will typically include the definition of ASCII together with descriptions of the format of the numbers and their locations in the file, their definitions as latitude and longitude, and the definition of their units as degrees” (CCSDS, 2002).

**Two possible (flawed) interpretations**
If Data Objects “interpreted using Representation Information” do directly “yield Information” then either:

- there is a conflation of entities: information content (latitude and longitude measurements) and its expression (the primary symbol structure: a table of numerals) conflated with the Information Object entity; or
- the primary symbol structure, the one expressing content, is not represented at all.

Both these interpretations will result in confusion and vulnerability if reflected in preservation systems designs.

**Restructuring OAIS Representation Information**
To have a correct, and sufficiently fine-grained representation of digital artifacts our model should include:

- **Digital objects** (the OAIS Data Objects): a lower level symbol structure in the form of a bit sequence
- **Primary expressions**: the table of numeral or the rendered image (in our examples)
- **Information content**: latitude and longitude measurement or the a set of feature a person can experience from a rendered image (in our examples)

**Representation information support only the performance of a digital object**

(1) The decoding of the primary symbol structure or the rendering of the image from the bits

**Conclusion**
Preservation actions like migrations can only be assessed if preservation models reflect a correct, complete, and sufficiently fine-grained representation of digital artifacts. Restructuring the OAIS account of Representation Information as described above will bring us closer to this.

But our analysis also reveals that there is still more work to be done: an account of how the primary (expressing, not encoding) symbolic expression (the performance of a digital object) is connected to content must also be provided. This is part of our ongoing research related to interpretive frames (Dubin et al., 2011).

**References**

**Acknowledgments**
The research reported here is being carried out at the Center for Informatics Research in Science and Scholarship (CIRSS) at the University of Illinois at Urbana-Champaign, Carole L. Palmer, Director. It is funded by the National Science Foundation, as part of the Data Conservancy, a multi-institutional NSF funded project (OCI/ITR-DataNet 0830976) hosted at Johns Hopkins University Sheridan Libraries.