

7th International Digital Curation Conference

December 2011

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

Center for Informatics Research in Science and Scholarship

Graduate School of Library and Information Science

University of Illinois at Urbana-Champaign

December 2011

The 7th International Digital Curation Conference will be taking place 5-7 December 2011 in Bristol. Please ensure you use the guidance in this template to produce your paper. Please submit your paper in one of the following formats: Microsoft Word (.doc), Open Office (.odt) or Rich Text (.rtf).



Poster Abstract

We describe two alternative interpretations of OAIS *Representation Information* (CCSDS, 2002), 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. This analysis is based on earlier work developing a framework for the application of significant properties to datasets (Sacchi et al, 2011).

As an illustration of how “*Data [Objects] interpreted using its Representation Information yields Information*” OAIS offers this example:

“*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).

We argue that in the example “ASCII table of number” is a data structure at a lower lever than suggested by the discussion. In the discussion there is a table (a *structure*) of numerals (*symbols*) that *expresses* latitude and longitude measurements (*information*). This table is *encoded* by an ASCII file — an OAIS Data Object — which is a lower level symbol structure (Sacchi et al., 2011).

This applies to all digital artifacts. Using a dataset example and a still image we show two cases:

- [1] Data Objects “interpreted using ... Representation Information” *do not* directly “yield Information”. *Data Objects* are symbol structures and interpretation via Representation Information “yields” (more precisely, is mapped to) not information, but another symbol structure – albeit a symbol structure that *expresses* information. In this case the OAIS definition of information as “*Any type of knowledge that can be exchanged*” does not hold.
- [2] If it *does* yield Information then either [a] there is a conflation of entities: knowledge as propositional content and its expression conflated with the *Information Object* entity; or [b] 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.

Representation Information, we argue, supports only the *encoding* and *decoding* of particular expressions, such as a table of numerals or a still image, into/from digital objects, such as an Excel file or a TIFF file, respectively.

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 expressions are connected to content must also be provided. This is part of our ongoing research related to *interpretive frames* (Dubin et al. 2011).

References

- [report] CCSDS, J. (2002). *Reference Model for an Open Archival Information System (OAIS)*. CCSDS 650.0-B-1, Blue Book.
- [proceedings] Dubin, D., Wickett, K. M., & Sacchi, S. (2011). Content, format, and interpretation. In B. T. Usdin (Ed.), *Proceedings of balisage: The markup conference*. Montreal, Quebec.
- [proceedings] Sacchi, S., Wickett, K. M., Renear, A. H., & Dubin, D. S. (2011). A Framework for Applying the Concept of Significant Properties to Datasets. *Proceedings of the 74th ASIS&T Annual Meeting on Bridging the Gulf: Communication and Information in Society, Technology, and Work*. New Orleans, Louisiana, October 9-12, 2011