

Data and Metadata Instantiation: Use Cases and a Conceptual Model

Presentation

Richard P. Smiraglia
 University of Wisconsin-Milwaukee, USA
 smiragli@uwm.edu

Keywords: data instantiation; metadata instantiation; DANS; clustering; disambiguation of groups; information retrieval; digital repositories

Abstract

Instantiation describes the phenomenon of variation in representation of information objects over time. Smiraglia (2008) describes it as the diatonic problem of both clustering and disambiguation of groups of what appear to be, but are not quite, iterations of the same object. Although the problem is well-known in bibliographic information retrieval (Smiraglia 2001), it also is well-documented among other kinds of information objects. Greenberg (2009) demonstrated instantiation among metadata records of evolutionary biologists, Coleman (2002) drew an analogy to instantiation among scientific models, and Smiraglia (2005) found instantiation among archival records of artifacts in a museum of archeology. As Greenberg points out (399), the problem is particularly acute in digital repositories where “automatic propagation, metadata inheritance, and value system adoption” contribute to a “lifecycle” that creates potentially ambiguous clusters.

Digital repositories are particularly susceptible to the problem of uncontrolled data and metadata instantiation because of the complex lifecycles of data deposit, use, and reuse. In repositories that require deposit of research data on a large scale, instantiation can become particularly acute. DANS (Data Archiving and Networked Services), a division of the Royal Netherlands Academy of the Arts and Sciences, is the “institute for permanent access to digital research resources” in The Netherlands (DANS 2017). The role of DANS is to encourage scholars to make their data accessible, interoperable and reusable, in a sustainable environment. In addition to serving as a host repository for tens of thousands of datasets, DANS also manages the NARCIS gateway to more than 160,000 datasets generated by Dutch scholars.

Recent research (Smiraglia and Park 2016) demonstrated one approach to a conceptual model of instantiation among open government data records, deriving core attributes “information object,” “expression,” “manifestation product type,” “actor,” “expression creation,” and “information carrier” from the FRBRoo ontology of bibliographic instantiation. The proposed presentation combines these and other FRBRoo attributes with the generations of lifecycle modeling identified by Greenberg, as applied to a series of use cases from DANS.

Works Cited

- Data Archiving and Networked Services (DANS). <https://dans.knaw.nl/en/about> accessed 12 June 2017.
- FRBRoo (Functional Requirements for Bibliographic Records—object oriented), an extension of the CIDOC-CRM. http://old.cidoc-crm.org/frbr_intro.html# accessed 12 June 2017.
- Coleman, Anita S. 2002. Scientific models as works. *Cataloging & classification quarterly* 33 no. 3/4:129-59
- Greenberg, Jane. 2009. Theoretical considerations of lifecycle modeling: An analysis of the dryad repository demonstrating automatic metadata propagation, inheritance, and value system adoption. *Cataloging & classification quarterly* 47, no.3:380-402.
- Smiraglia, Richard P. 2001. *The nature of a work: Implications for the organization of knowledge*. Lanham, Md.: Scarecrow Press.

- Smiraglia, Richard P. 2005. Content metadata: An analysis of Etruscan Artifacts in a museum of archeology. *Cataloging & classification quarterly* 40 no. 3/4:135-51.
- Smiraglia, RICHARD P. 2008. A meta-analysis of instantiation as a phenomenon of information objects. *Culture del testo e del documento* 9, no. 25:5-25.
- Smiraglia, Richard P. and Hyoungjoo Park. 2016. Using Korean Open Government Data for Data Curation and Data Integration. Presentation, International Conference on Dublin Core and Metadata Applications, DC-2016, Copenhagen, Denmark. <http://dcevents.dublincore.org/IntConf/dc-2016/paper/view/447/509> Accessed 12 June 2017.