

Richard P. Smiraglia, Professor
Palmer School of Library and Information Science
Long Island University
720 Northern Blvd.
Brookville NY 11548 USA
(516) 299-2174 voice
(516) 299-4168 fax
(215) 266-5072 mobile
Richard.Smiraglia@liu.edu

INSTANTIATION: EMPIRICAL EMERGENCE OF A GLOBAL PHENOMENON

Abstract: Knowledge extraction for resource-description ontology via empirical research yields better results than the traditional rational construction of schema. An example is the case of instantiation. In this paper, empirical methods for the derivation of resource content and the description of instantiation are presented, thus posing one possible model for the empirical derivation of categories, descriptors, and schemas. Instantiation will be demonstrated as one example of a semantically interoperable model for the integration of scientific and cultural information. This model has been derived empirically in several studies, which are summarized here. These include epistemological analysis of the concept of instantiation (Smiraglia 2005), evidence from several studies of bibliographic works (reported by Smiraglia 2001), case-study evidence from Etruscan artifacts at The University of Pennsylvania Museum of Archeology and Anthropology (Smiraglia 2004a and b), and The Archives of The Class of 1942 of The United States Merchant Marine Academy (Smiraglia 2006). A model for mapping instantiation networks is demonstrated.

1. Background

Global resource sharing requires semantic interoperability on several levels, from metadata structures to the representation of intellectual content. However, interoperability is hampered by a lack of empirical evidence supporting the development of data structures. The post-modern emphasis on domain-specific ontology, which is widely seen as an improvement in knowledge organization, can lead to a potential form of “Balkanization” (that is, isolation) in resource description. Repository- and resource-specific data-structures, generated in-house by rational or pragmatic means, often fail to acknowledge the universality of function that might be common among resources. Knowledge extraction for ontology via empirical research yields better results. An example is the case of instantiation. When is a terracotta hut urn like a sailor’s deck-log like a best-selling novel? When its representations are multiply instantiated within a repository, but the instantiations are not distinguishable for retrieval purposes.

Metadata for resource description have traditionally been derived rationally—one might also say qualitatively—generated inductively from case to case rather than derived empirically as the result of research. Metadata are categorical descriptors of information resources, often used as alphabetico-classed segments of thesauric-faceted strings for information retrieval. That is, resource-linkage is provided through the semantic application of content metadata with the dual purpose of collocation and disambiguation. Furthermore, specific metadata structures provide semantic linkage among resource descriptions. Hjørland, applying activity theory (2003, 98) to the act of naming informing objects (documents, artifacts, records, and

their content, to be specific) says that “naming” is the action of facilitating use. Metadata, essentially a set of “names,” cannot be neutral because the names influence the activity of facilitating (or obfuscating) use. Metadata schemas, rationally deduced, predetermine the potential use of intellectual content by limiting its retrieval. Greenberg (2005, 30) also suggests that data modeling, a way of typing objects or naming them to facilitate their use must be empirically-driven. Knowledge extraction should be based on empirical observation of the resource itself.

“Instantiation” is the phenomenon first denoted empirically by research into bibliographic ‘works.’ Specifically, an instantiation of a work exists whenever the work is manifest in physical form (in a book, for example). A problem arises when multiple instantiations of a work (several editions, translations, etc.) exist and their descriptions must be collocated (or linked) in a retrieval system with sufficient information to assist in the selection of the instantiation of interest to a searcher. Similarly, unique artifacts can be represented by metadata or images (called representations), which can exist in multiple instantiations (a photographic negative, a print, its digital descendent, etc.). The same is true of the representations of archival documents, which might exist in paper photocopies (original and carbon), digital images, and so forth.

2. Empirical Research

Several studies have demonstrated consistent theoretical parameters for the concept of instantiation, across resource and repository types. Epistemological analysis of the concept of instantiation provided a background for meta-analysis of the phenomenon of instantiation (Smiraglia 2005). Empirical evidence from several studies of bibliographic works (reported by Smiraglia 2001) were analyzed together with case-study evidence from Etruscan artifacts at The University of Pennsylvania Museum of Archeology and Anthropology (Smiraglia 2004a and b), and from the ongoing project to digitize the archives of The Class of 1942 of The United States Merchant Marine Academy (Smiraglia 2006).

The analysis of the Etruscan artifacts demonstrated the presence of both representations (images, models, etc.) and metadata sets (descriptive data) for each artifact. The term “content genealogy” was used to describe the succession of representations that occur along a chronological continuum. Rich examples included:

- Museum database contains semantic content of object-card embedded in new dataset = successive relationship ; amplification relationship
- Museum images multiply instantiated (negative, slide, print, digitized, etc.) appear in print or on web = simultaneous ; successive relationship
- Artifacts publicized at acquisition have larger instantiation sets = cultural catalyst

Each artifact had an equivalent metadata set in-house, which included a core ‘object entity’ made up of: Type, Material, Culture, Source, Collector, and Date Acquired. To the object entity dataset was appended many work- or activity-based descriptors that varied according to museum function (field notes, accession ledgers, object records, conservation records and reports, and online catalog records). Object-type metadata sets included common descriptors, but activity-based descriptors differed from division to division depending on the work that had taken place.

Analysis of the merchant marine archives demonstrated a collection rich with letters, envelopes, binders, photographs, ship's deck-logs, time-sheets, scholarship applications, and so forth. The archives of The Class of 1942, compiled by the enterprising graduates of this class, include large instantiation networks of the documents pertaining to their service. These sets include: photocopies, carbon copies, digitized scans of postcards containing photographs, scans of photos, photos alongside digitized scans of them, and documents together with their carbon copies and digitized scans of the originals.

Multiplicity among informing objects is universal and the analysis of instantiation has demonstrated this. It also demonstrates the usefulness of empirical strategies for generating categories and their descriptions. Interestingly, tabulation of instantiation networks revealed that the terms represented not so much categories as properties. The categories were not mutually exclusive, but could appear together in the same instantiation. That is, just as a tree may be deciduous, so a translation may have a successive edition or appear in a commentary or both. This suggests that what we are dealing with is not so much taxonomy as typology.

The typologies of instantiation, when analyzed together lend support to Hjørlands idea that activity theory can proscribe the categorizing activity of knowledge organization. We denote categories so as to assign information objects spatial loci within the acknowledged schema. The terms in these typologies, empirically derived, represent the properties of instantiation in three contexts, and yet they consistently demonstrate the epistemological properties of mutation and derivation (see Smiraglia 2002b). Derivation denotes types or properties of instantiation in which intellectual content is unaltered; mutation denotes types or properties of instantiation in which intellectual content has been altered semantically or ideationally. Mutations have been observed to occur in bibliographic and artifactual domains, and derivations have been observed to occur in bibliographic, archival, and artifactual domains.

Other consistent elements across these studies of instantiation include: 1) the concept of "canonicity" as a predictor of instantiation; 2) the influence of time as a predictor of the degree of instantiation; 3) transitive relations demonstrated at points of change in semantic or ideational content as predictors of the type of instantiation; and 4) a continued association of the incidence of instantiation with Lotka's law.

3. Mapping Instantiation

One can map instantiation networks logically in the following manner. For every information object O_i the possibility of instantiation is present. The historic action of some catalytic influence results in the presence of an instantiation I_i . The simplest set is the set (an edition of a book, a photocopy of a document, a photograph of an artifact):

$$O_i \rightarrow I_i$$

A set of instantiations—an instantiation network--might run from I_{i-n} . A common set is the set (many editions of a book appear over time, many photocopies of a document are produced, or photographs of an artifact are produced repeatedly from an original negative image):

$$O_i \rightarrow I_{i-n}$$

Catalytic influence, the key to instantiation, is a culturally sensitive phenomenon, resulting from some demand close to the object of origin. These influences occur at more widely

spaced points over time as the trajectory recedes temporally. A work that is popular in the time of its origin might appear in multiple editions, then recede for centuries until a cultural shift “rediscovered” it, and generates new sets of instantiations. This yields instantiation nodes within the set, from which instantiation networks can also proceed:

$$O_i \rightarrow I_{i-n} I_{ii-n} I_{iii-n}$$

where each I_i is a new instantiation node. At each node n is the sum of the i s (hence the extent of the instantiation network), and I_x is the representative of a node of instantiation.

In this paper, empirical methods for the derivation of resource content and the description of instantiation are described, thus posing one possible model for the empirical derivation of categories, descriptors, and schemas. Instantiation will be demonstrated as one example of a semantically interoperable model for the integration of scientific and cultural information. This model has been derived empirically.

Works Cited

Greenberg, Jane. 2005. Understanding metadata and metadata schemes. *Cataloging & classification quarterly* 40n3/4: 17-36.

Hjørland, Birger. 2003. Fundamentals of knowledge organization. *Knowledge organization* 30: 87-111.

International Federation of Libray Associations. 1998. *Functional Requirements for Bibliographic Records*. Munich: K.G. Saur, 1998. Available <http://www.ifla.org/VII/s13/frbr/frbr.htm> or <http://www.ifla.org/VII/s13/frbr/frbr.pdf>

Smiraglia, Richard P. 2001. *The nature of a work: implications for the organization of knowledge*. Lanham, MD: Scarecrow.

Smiraglia, Richard P. 2004a. “Knowledge sharing and content genealogy: extending the “works” model as a metaphor for non-documentary artifacts with case studies of Etruscan artifacts.” In McIlwaine, Ia C., ed. *Knowledge organization and the global information society; Proceedings of the eighth international ISKO conference 13-16 July London UK*. Advances in knowledge organization v. 9. Würzburg: Ergon Verlag, pp. 309-14.

Smiraglia, Richard P. 2004b. “Content genealogy and the development of metadata for resource description.” *Cataloging & classification quarterly* (at press).

Smiraglia, Richard P. 2005. "Instantiation: Toward a theory." In Data, information, and knowledge in a networked world: Proceedings of the Canadian Association for Information Science annual conference June 2-4 2005, ed. Liwen Vaughan. <http://www.caais-acsi.ca/search.asp?year=2005>.

Smiraglia, Richard P. 2006. “Empiricism as the Basis for Metadata Categorisation: Expanding the Case for Instantiation with Archival Documents.” In Budin, Gerhard, Christian Swertz and Konstantin Mitgutsch eds. *Knowledge organization for a global learning society: Proceedings of the ninth international ISKO conference 4-7 July Vienna, Austria*. Advances in knowledge organization v. 10. Würzburg: Ergon Verlag, pp. 383-88.