

Presentation
**Using Korean Open Government Data for Data Curation and
Data Integration**

Richard Smiraglia
University of Wisconsin-
Milwaukee, U.S.A.
smiragli@uwm.edu

Hyoungjoo Park
University of Wisconsin-
Milwaukee, U.S.A.
park32@uwm.edu

Abstract

This presentation addresses cultural heritage data-sharing practices through the use of Republic of Korea open government data for data-curation and data integration. Data curation enables data-sharing throughout the data management life cycle to create new value for new user needs. Previous studies for cultural heritage data integration have been conducted with the mediation between metadata and ontology. Examples are ontology-based metadata integration in the cultural heritage domain with mediation between Dublin Core (DC) and the meta-level ontology known as the CIDOC CRM (International Committee for Documentation - Conceptual Reference Model) (Stasinopoulou et al. 2007), DCMI type vocabulary and the CIDOC CRM (Kakali et al. 2007), DC metadata and the CIDOC CRM in cultural heritage digital object collections (Koutsomitropoulos, Solomou and Papatheodorou 2009), and between archival metadata such as Encoded Archival Description (EAD) and the CIDOC CRM (Bountouri and Gergatsoulis 2011). A gap remaining from prior studies is that cultural heritage data integration has not been actively studied with an emphasis on knowledge organization and data curation using open government data.

Our research employed a visualization phase, in which we used domain analytical techniques to better understand the contents of the population of 375 library-related open government cultural heritage data available at the Korean Open Government Website (<http://data.go.kr/>). Researchers translated all records from Korean to English. Data were in unstructured and in heterogeneous formats such as file formats, data formats and or web addresses.

For data curation and integration, we employed the meta-level ontology known as the CIDOC-CRM, which we applied qualitatively to small sets of carefully selected records. This phase was based on an earlier project using a different data-set (Park and Smiraglia 2014), in which cultural disparities between Korean data and the CRM were detected and resolved. Visual mappings are conducted by using the mapped Korean open government data which were in unstructured and heterogeneous formats by using CIDOC CRM version 6.2. The mappings were simple and straight-forward.

To map instantiation of records, which is required for data integration, we used FRBRoo (Functional Requirements for Bibliographic Records – object oriented), an extension of the CIDOC CRM, to map the instantiation of data records in a typical data-sharing scenario. Then, equivalent mapping processes were comparatively tested with visualizations to demonstrate the effective harmonization between the CIDOC CRM and FRBRoo, which enables the integration of metadata and data curation from unstructured and heterogeneous formats. This presentation may contribute to the cross- or meta-institutional integration of curation across institutional boundaries in cultural heritages as an imperative for cultural synergy and the role of information institutions (Smiraglia 2014) with metadata integration.

References

- Bountouri, Lina, and Manolis Gergatsoulis. 2011. "The semantic mapping of archival metadata to the CIDOC CRM ontology." *Journal of Archival Organization* 9: 174-207.
- Kakali, Constantia, Irene Lourdi, Thomais Stasinopoulou, Lina Bountouri, Christos Papatheodorou, Martin Doerr, and Manolis Gergatsoulis. 2007. "Integrating Dublin Core Metadata for cultural heritage collections Using ontologies." *DCMI International Conference on Dublin Core and Metadata Applications*. Singapore. 128-139. Accessed July 13, 2016. <http://dcpapers.dublincore.org/index.php/pubs/article/view/871>.
- Koutsomitropoulos, Dimitrios A., Georgia D. Solomou, and Theodore S. Papatheodorou. 2009. "Metadata and semantics in digital object collections: A case-study on CIDOC-CRM and Dublin Core and a prototype implementation." *Journal of Digital Information* 10 (6).
- Park, Hyoungjoo, and Richard P. Smiraglia. 2014. "Enhancing data curation of cultural heritage for information sharing: A case study using open government data." Edited by S. Closs, R. Studer, E. Garoufallou and MA Sicilia. *Metadata and Semantics Research*. Karlsruhe: Springer International Publishing. 95-106. Accessed July 10, 2016. doi:10.1007/978-3-319-13674-5_10.
- Smiraglia, Richard P. 2014. *Cultural synergy in information institutions*. New York: Springer.
- Stasinopoulou, Thomais, Lina Bountouri, Constantia Kakali, Irene Lourdi, Christos Papatheodorou, Martin Doerr, and Manolis Gergatsoulis. 2007. "Ontology-based metadata integration in the cultural heritage domain." Edited by Dion Hoe-Lian Goh, Tru Hoang Cao, Ingeborg Torvik Sølvsberg and Edie Rasmussen. *10th International Conference on Asian Digital Libraries*. Hanoi: Springer Berlin Heidelberg. 165-175. Accessed July 13, 2016. doi:10.1007/978-3-540-77094-7_25.