

Knowledge Organization and Cultural Heritage in the Semantic Web – A Review of a Conference and a Special Journal Issue of *JLIS*

Marcia Lei Zeng <mzeng_at_kent_dot_edu>, Kent State University, Kent, Ohio, USA
Sophy Shu-Jiun Chen <sophy_at_sinica_dot_edu_dot_tw>, Academia Sinica, Taiwan

Abstract

Review of the “Knowledge Organization and Cultural Heritage: Perspectives of the Semantic Web” conference held at the *Academia Sinica Center for Digital Cultures* in Taipei on June 2, 2016 and a special journal issue of academic papers related to the conference. Speakers came from mainland China and Taiwan across the Taiwan Strait, as well as the United States. Topics included methods and theories of Linked Open Data (LOD), the modeling of ontologies and knowledge bases, the historical and structural review of knowledge organization and representation, practical approaches used in exploring cultural assets, and the efforts of standards development for preservation and management of digital data. Speakers were encouraged to submit full papers to the peer-reviewed *Journal of Library and Information Science* (ISSN 0363-3640). These led to the publication of a whole special issue of the journal (vol. 43/no.1) in April 2017.

Recently, Digital Humanities have garnered much attention all over the world. Concrete results have been produced by the collaboration between humanistic studies and the interdisciplinary research of semantic technology and information science. Most of these studies rely on the deep excavation of cultural assets, and the complete reorganization and cross-border interlinking of information resources held by libraries, archives, museums (LAMs), digital centers, data repositories, and more. 1

In this rapidly evolving field of cross-disciplinary studies, two imperatives have emerged that require further research and discussion: (1) How to construct Knowledge Organization Systems (KOS) following current Semantic Web standards and best practice guidelines? This involves combining various methods applicable to digital humanities, and conducting studies aiming to promote, utilize, and preserve cultural heritages along with the digital datasets based on and created for them; (2) How can we structure and express the data in the digital assets so they can be handled and understood by both ordinary users and machines? This question is raised because the creation, linking, enhancing, and reusing of a semantic-rich dataset depends on the usability, flexibility, and reliability of its knowledge organization structure and encoding schema. 2

To explore in greater depth these two essential needs which have come to light in the field in recent years, the *Academia Sinica Center for Digital Cultures* organized the “Knowledge Organization and Cultural Heritage: Perspectives of the Semantic Web” conference on June 2, 2016 in Taipei [ASCDC 2017]. Speakers came from mainland China and Taiwan across the Taiwan Strait, as well as the United States. Topics included methods and theories of Linked Open Data (LOD), the modeling of ontologies and knowledge bases, the historical and structural review of knowledge organization and representation, practical approaches used in exploring cultural assets, and the efforts of standards development for preservation and management of digital data. All topics were presented from different perspectives and were extremely well received by the attendees. As a result, after the conference, speakers were encouraged to submit full papers to the peer-reviewed *Journal of Library and Information Science* (ISSN 0363-3640). These papers led to the publication of a whole special issue of the journal (vol. 43/no.1) in April 2017 [Zeng and Chen 2017]. Each full article (in Chinese) includes a long English abstract, and all are openly available on the Web from the journal’s website at 3

The nine papers presented in this special issue of the *Journal of Library and Information Science* put forward answers to the aforementioned challenges that LAMs and other cultural heritage organizations encounter. The reported solutions utilized semantic technologies in conjunction with knowledge organization methods in order to progress from digitization to datafication. Datafication is critical because it makes possible production of data that is not just machine-readable but truly machine-understandable, -processable, and -actionable. While these are challenging topics, they hold the most up-to-date concepts and breakthroughs of interest to humanities researchers, information scientists, and Semantic Web developers, in addition to practical answers for the pressing needs of LAMs, digital centers, and repositories. The papers are presented in four broad areas:

1. To address the problem of reusability of structured data, the paper titled “Reuse of Structured Data: Semantics, Linkage, and Realization” by researchers from the Institute of Information Science at Academia Sinica, Taiwan (Huang, Lee, and Chuang) presents in detail the five cross-domain knowledge bases and seven special domain knowledge bases currently using LOD approaches. The paper further reports the results of data quality analysis and the relations between metadata and data provenance, providing a lightweight ontology voc4odw for describing data reuse and provenance. The website data.odw.tw stores and converts the collection of metadata into semantically-structured linked data, reflecting the steps taken from original cataloging data to semantic linked data segmentation.
2. Knowledge organization structures and semantic models are used in cultural asset preservation, management, research utilization, and practical applications. With the purpose of understanding the current development and application of knowledge organizations, this special journal issue included four articles demonstrating the recent developments and applications of knowledge organization systems (KOS).
 - Cuijuan Xia’s “Building a Digital Humanities Platform by Using Linked Open Data Services” describes how the Shanghai library has used LOD on a large scale. Through reforming the traditional methods of digital resource collections, the team established a series of independent yet interlinked knowledge bases. The design process and the methods for fully integrating heterogeneous literature resources to form the two literature knowledge bases (“Sheng Xuanhuai Archives Knowledge Base of Shanghai Library” and “Genealogy Knowledge Service Platform of Shanghai Library”) are explained thoroughly, demonstrating the new role and innovative usage of historical archives in the Semantic Web era.
 - “A Study of Linked Data for Digital Collections: A Case of the Painter Chen Cheng-Po” by Shu-Jiun Chen from Academia Sinica, focuses on digital collections, particularly person-centered digitized materials. The project aims at presenting a more flexible search function and enabling exploration in the context of a group of human characters centered around the artist Chen Cheng-Po, responding to complex questions based on semantically modeled and visualized relationship networks. The exemplary contribution of this research lies in the research method, which allows machines to handle and understand large amounts of digital humanities resources. As a result, the research datasets can be easily used by users and re-organized when building their own new research, thus enhancing the effectiveness of data use.
 - Authors from Wuhan University and Dunhuang Research Institute (Wang, Liu, and Xia) presented “Design and Implementation of Deep Semantic Indexing on Digital Cultural Heritage Images.” Using one of the Dunhuang cave’s “Nine-Colored Deer” paintings as an experiment, the paper expresses the deep image indexing method through analyzing semantic features and themes of the images. It also reviews existing metadata schemes and ontologies for cultural heritage objects. Established on the concept of deep semantic indexing and its basic requirements, a workflow was designed and three semantic indexing models were constructed: the macroscopic concept model, information hierarchy model, and structured image annotation model. This kind of deep semantic indexing theory and the application of image information organization theory are of great significance to the development of image information organization methodology and digital humanistic research.

- Using ontological and Linked Data approaches, Fu and Ke's "Ontology-Based System for Librarianship Development - A Case Study of Professor Chen-Ku Wang" designs a knowledge ontology through analyzing Professor Chen-Ku Wang's lifelong academic achievements and contributions. The ontology was used to describe his scholarly history. The aim of the research was to build a website describing the history of scholars in the field of library and information science in Taiwan. In addition to browsing, searching, and linking functions, a user can also access or download the ontology and knowledge base content through SPARQL queries, allowing the website to become a provider of Linked Data. The paper provides the results of the website's efficiency test analysis.
3. Knowledge organization and standardized value vocabularies are closely related. How can libraries (which have a tradition of implementing knowledge organization systems) keep up with the 21st century Web, especially embracing the wave of new ontology development? Three papers in this special issue reported national-level efforts on the development of classification, thesauri, and metadata standards demonstrating how this can be done.
- First, Wei Fan's article, "Towards Open, Semantic, and Linked *Chinese Classified Thesaurus*" presents the development of the *Chinese Classified Thesaurus* (CCT) seen from three milestones: open, semantic, and linked. The development stages of CCT since 1994, as well as the research and practice of advanced information retrieval languages are fully explained in the context of knowledge organization in the field. It concluded that the Semantic Web presents new opportunities for libraries and information organizations.
 - The needs for knowledge organization at the Internet age are apparent and beyond the scope of library collections and published literature. Cultural heritage institutions and museums heading towards digitization have become the new norm. However, in terms of information standards for cultural heritage assets, new norms are not yet established. "Methodology for the *Faceted Thesaurus of Chinese Cultural Heritage: An Initial Discussion*" by Ming-Yu Huang from University of Science & Technology Beijing, uses China's *Faceted Thesaurus of Chinese Cultural Heritage* (FTCCH) and the USA Getty Research Institute (GRI)'s *Art & Architecture Thesaurus* as a base, detailing the background and process of FTCCH, aiming at developing the concepts and vocabularies for Chinese cultural heritages in this needed thesaurus. The results at the current stage attempt to establish preliminary methodology and the initial structure of FTCCH which led to a significant product of the FTCCH.
 - Related to FTCCH's development and implementation, Peking University's Long Xiao et. al. shared "A Cross-disciplinary and Fundamental Study of Digital Humanities: Taking 'Metadata Standards for Digital Cultural Heritage Data' as an Example." The article introduces a metadata standard for digitized cultural heritage data usage in China. A metadata standard has been developed through investigation and analysis, field research, general framework modeling, standardized designs, and other steps. As a component fitting in the overall structure of the "Digital Preservation of Cultural Heritage" project in China, the paper clarifies the necessity and function of a standard norm, while exploring how humanities and technology should cooperate and grow together.
4. All the latest developments in the digital age and knowledge organization require a theoretical perspective to aid in understanding. Thus, the last paper in this issue is Mei Mei Wu's "Some Thoughts on Knowledge Organization in the Web Era," which posits that in the pre-Internet paper age, a single informational item is like an isolated island in terms of knowledge organization, whereas in the Internet age there is a drastic change. Traditional methods of identification and searching are replaced by linked and multi-oriented methods which allow clues and knowledge to be reconstructed, organized, and interpreted to generate new knowledge. In this Internet age, knowledge organization contributes to how humans comprehend knowledge. The paper further proposes a kind of education for future knowledge organization professionals.

In conclusion, this journal issue gathers the latest research developments of Taiwan and Mainland China across the Taiwan Strait, exploring multiple aspects via collective efforts, studies, and examples. The conference and this special issue intend to facilitate cross-disciplinary research and discussion between knowledge organization professionals and Semantic Web developers, as well as to contribute to the growing body of digital humanities research by sharing experiences and promoting interdisciplinary communication. The current role and function of knowledge organization is revisited, and implementation requirements articulated. As demonstrated by the contributors of this conference and special journal issue, for those researchers in need of historical data, the information resources provided by LAMs and digital centers have extraordinary value. Knowledge organization theories, models, systems, and practical approaches are fundamentals to: enabling the deep excavation of cultural assets, transforming semi-structured and unstructured data into structured and machine-processable data, contextualizing and linking of existing structured data across data silos, and enabling one-to-many usages of LAM data in supporting digital humanities.

This collection of papers from many experts came from different perspectives, but the central idea embodied is clear: from the Semantic Web perspective, knowledge organization systems and knowledge organization methods are in the midst of a paradigm shift. Opportunities brought by semantic technologies are driving us to redefine the definition of knowledge organization, encouraging us to re-examine the changes, potentials, and breakthroughs in information science.

Works Cited

ASCDC 2017 Academia Sinica Center for Digital Cultures (ASCDC). "Knowledge Organization and Cultural Heritage: Perspectives of the Semantic Web." 2017: <http://seminar2016-semanticweb.ascdc.tw/index.html> [Accessed 20 February 2018].

Zeng and Chen 2017 Zeng, Marcia Lei, and Shu-Jiun Chen (Eds.). "Special Issue," *Journal of Library and Information Science* 圖書館學與資訊科學 (ISSN 0363-3640) 43.1 (April 2017): <http://jlis.glis.ntnu.edu.tw/ojs/index.php/jlis/issue/view/89> [Accessed 20 February 2018].



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.