Book Review

Edited by Hanne Albrechtsen

Book Review Editor

Handbook of Metadata, Semantics and Ontologies. Edited by Miguel-Angel Sicilia. Singapore: World Scientific Publishing Co. 2014. 570 p. ISBN 978-981-283-629-8, price: 155 US$

Never would I have ever thought to write a review about a book from a computer scientist. However the title provoked me, I wanted to know what "they" know about our field of knowledge organization. And here it goes.

For this handbook of 570 pages, the compiler, Miguel-Angel Sicilia from the University of Alcala, Spain, assembled 20 contributions from 36 authors of 13 countries (Austria, Belgium, Brazil, Chile, Ecuador, England, Germany, Greece, Ireland, Italy, Mexico, Spain, and the USA). He divided this collection into 4 sections: (I) four (plus one) contributions on metadata, (II) four on ontologies, (III) eight on applications in special domains, and (IV) three on technologies for interoperability, metadata extraction and metadata and ontology storage.

Metadata—an expression coined already in 1968 by Philip Bagley (Wikipedia) is being understood as data about data, and according to Sicilia is "a generic term of any kind of meta-information." With the creation of the Web, "metadata has been brought to the hearth of the architecture of cyberspace." The expression media for metadata today are XML and RDF. In the language of KO experts, metadata may be understood as concepts: for information scientists as descriptors; for taxonomists as taxa; and for librarians as subject headings. In the meantime the latter adopted already 'metadata,' apparently for its simplicity and possible collaboration with computer people, the same happened—but in a much wider sense—with the term 'ontologies' used for classification systems, taxonomies, thesauri, vocabularies and the like, together they are termed by computer people 'semantics.' Although absurd, as 'ontology' is a subdiscipline of philosophy, this expression in the plural has made in the meantime a triumphal procession into all domains working towards organizing knowledge elements into systems—with the exception of the taxonomists.

With his first contribution, Chapter I.1, on "Metadata Research: Making Digital Resources useful again," A.-M. Sicilia set the frame for the handbook and called 'metadata' a research discipline, and by trying to define it more extensively, referred to different forms of metadata today. Incidentally he deplores "the lack of a pertinent professional society or organization"—apparently he is not as yet aware of the existence of ISKO, the International Society for Knowledge Organization. Finally he refers to the newly started "Linked Data ... as a simple but extremely powerful concept (sic!) that overcomes limitations of microdata and XML-based harvested collections." They provide "the fundamental properties to build a coherent global system of metadata in the coming years ... they relate useful information to other information ... with RDF as implementation technology." (For an example of this possibility see the reference at the end.)

Chapter I.2 deals with "Metadata Typology and Metadata Uses;" by Eva Mendez from Madrid and Seth van Hooland from Brussels. They provide an overview of how metadata can be interpreted differently and give a typology of metadata schemes and elements and how they can be deployed by user communities. The authors of Chapter I.3, M.D. Lytras from Greece, A.-M. Sicilia (Spain) and C. Cechinel from Brazil wrote about "The Value and Cost of Metadata" considering as metadata only those data which are in digital form and which are a subject to be used through communication technologies, managed through software to enhance information retrieval as an alternative or complement to text-based IR algorithms and formal bibliographic records. Their value and costs are dependent on the functions which vary in the many different sciences using them; they may increase usefulness and create benefits. Table 1 shows examples of metadata systems and their typical characteristics with their quality and evaluation policy as well as sources of value. In Table 2 some metadata cost examples are given.

Chapter I.4 by X. Ochoa from Ecuador is about "Metadata Quality," presenting the main theoretical approaches to the definition of metadata quality, so far poorly researched because of inherent complexity and subjectiveness. A selection of strategies and tools to evaluate the quality of metadata is given.

Chapter I.5 on "Ontologies in Systems Theory" by E. Curras from Madrid should belong in the next section. It aims at applying systems theory to ontologies, which are

to be understood as "complex, conceptual, empirical classification systems," where the "principal and primary node is the word."

Section II, as said before, to be meant on ontologies, starts with Chapter II.1 by L. Papaleo from Italy and gives an "Introduction to XML and its Applications." The abstract says: "Extensible Markup Language (XML) is a meta language for defining new languages. Its impact on the modern and emerging web technologies has been and will be incredible." I must confess, this article was most informative and helpful for me, also as it contains the explanation of very many respective abbreviations. It explains what XML is, its origin in SGML and the application of the latter in HTML, further headings are: The birth of XML, Syntax of XML documents, Name spaces, Structuring XML documents, Element declarations, XML schemas, Rendering XML documents via CSS (Cascading Style Sheet), Transforming and rendering XML documents.

Chapter II.2 by S. Arroyo from Spain and K. Siorpaes from Austria in "Ontologies and Ontology Languages" give a general overview on approaches to model metadata by controlled vocabularies, taxonomies, thesauri and ontologies. These systems are explained in order to understand their later possible change into a form that is interpretable for the computer. For this, so-called ontology languages are necessary, the most relevant ones are: RDF(S) Resource Description Framework, a general purpose language to represent information about resources in the Web; OWL, Web Ontology Language, which exists in three versions: WSML, Web Service Modeling Language, which has four versions; SKOS, Simple Knowledge Organization System which stands for "all kinds of controlled vocabularies describing the minimal set of classes and properties that are necessary to express knowledge in simple structures such as taxonomies" and finally OIL built on RDFS, but superseded by the OWL family of languages.

Chapter II.3 by P. G. Picazo and J. Tramulas, both from Spain, deals with "Topic Maps." These have already become an ISO standard, reflecting the importance of this tool to describe and organize digital information resources. Chapter II.4 has three authors from Germany: T. Bürger, E. Simperl and Ch. Tempich: "Methodologies for the Creation of Semantic Data." They provide an overview of knowledge engineering methodologies for knowledge-based systems as well as their tool environments supporting their application. Industrial take-ups of these semantic technologies seem possible.

For Section III with applications of metadata and ontologies in eight special domains M. G. Sicilia, the compiler, selected some rather unusual but also very voluminous domains. Here we just mention the authors and their titles: Chapt.III.1: M. E. Prieto. Mendez (Spain), V. H. Menendez Dominguez (Mexico) and Ch. L. Vidal Castro (Chile): "Metadata and Ontologies in E-Learning." Chapt. III.2: G. Colombo, D. Merico, M. Gündel (Italy): "Metadata and Ontologies for Health." Chapt. III.3: N. Palavitsinis, N. Manouselis (Greece): "Agricultural Knowledge Organization Systems: An Analysis of an Indicative Sample." Chapt. III.4: E. Blanco (Spain): "Metadata and Ontologies for Bioinformatics." Chapt.III.5: F. Sarton, S. Bandini (Italy): "Metadata and Ontologies for Mechanical Objects: Design and Manufacturing." Chapt.III.6: L. Santos-Santos, T. Aguado-Goméz, (Spain): "Metadata and Ontologies for Emergency Management." Chapt.III.7: D. Kanellopoulos (Greece): "Metadata and Ontologies for Tourism." Chapt.III.8: T. Bürger (Germany), M. Hausenblas (Ireland): "Metadata Standards and Ontologies for Multimedia Content." In almost all of these articles it was astonishing for me that the work of information scientists in elaborating thesauri and classification systems had apparently been very useful instruments for the work of these authors. In the case of Chapt.III.2 "Health," the authors gratefully acknowledged the work towards a *Unified Medical Language System (UMLS)*, especially also for the definitions of their terms!

The three chapters of the following Section IV appeared to me as most valuable for every newcomer into this new knowledge-field for information scientists, taxonomists and librarians. Indeed, their authors with their references reveal a familiarity with "our business." Chapt.IV.1: R. Eito-Brun, (Spain): Technologies for Metadata Integration and Interoperability. Chapt.IV.2: K. Golub, H. Muller, E. Tonkin (England): Technologies for Metadata Extraction. Chapt.IV.3: M. Parmelee, L. Obrst (USA): Technologies for Metadata and Ontology Storage.

Doubtlessly, this Handbook of Prof. Sicilia is a laudable undertaking and must have been really hard work. There was apparently no time left for a subject index or a list of abbreviations with their interpretation. Laudable also is the number of references to each article. I took the "pain" to count all of them. If I am not mistaken, I found altogether 963! A really rich little bibliography of this field!

The volume was written by and for computer science people (with two exceptions). In some cases they cited at the end the grants by which they were able to do the investigation. Not only with regard to the last three chapters but in some way also for the others, it might serve our community quite well. Unfortunately the whole undertaking came to appear only in January 2014 and although I ordered it already in January, I received the review copy in April. And it must have taken five years to assemble, write and check all the contributions, visible in the year indications of the references, the latest ones are from 2009! Therefore a number of important developments since then could not be included.

It seems to me, a new way of possible acquaintance and collaboration between knowledge organization and knowledge engineering communities ought to be taken into consideration, perhaps by textbooks written from each side for the other side. The three last chapters of the Handbook would pave the way.

Reference

DeLuca, Ernesto William and Dahlberg, Ingetraut. 2014. Including knowledge domains from the ICC into the multilingual lexical linked data cloud. In Babik, Wieslaw, ed., *Knowledge organization in the 21st century: between historical patterns and future prospects: Proceedings of the Thirteenth International ISKO Conference 19-22 May 2014 Kraków, Poland.* Advances in knowledge organization 14. Würzburg, Germany: Ergon, pp. 258-65

Ingetraut Dahlberg
Bad König, May 29, 2014