Book review: Seth van Hooland, Ruben Verborgh. Linked data for libraries, archives and museums : how to clean, link and publish your metadata. London : Facet Publications, 2014. xvii, 254 pages. ISBN: 9781856049641

Authors Seth van Hooland and Ruben Vergogh are metadata specialists. The former is an assistant professor at the Universite libre de Bruxelles (ULB), while the latter is a researcher in semantic hypermedia at Ghent University – iMinds, Belgium.

The introduction is very helpful, describing the structure of the book by outlining the goals, audience, concepts and skills of each chapter. The authors ask whether linked data is “the kingdom of structured data to come or an irritating buzzword which we all will have forgotten in a few years?” They also state that “The ambition of this handbook is to bring a sense of pragmatism to the debate.” Each of the five core chapters (modelling, cleaning, reconciling, enriching and publishing) leads to a case study with metadata from institutions around the world. At the beginning of each chapter is a list of learning outcomes. There is also a useful glossary at the beginning of the book. The authors’ intention is for the chapters and case studies to stand on their own and so they can be read individually. Overall, I found the book to be well organised and clearly set out.

Modelling provides an overview of four data models: tabular data, relational model, meta-markup languages and linked data. This chapter makes it clear that every model has been developed for a specific use.

The chapter on cleaning claims that “all metadata is dirty, but you can do something about it.” They go on to explain that “recurrent metadata quality issues such as duplicate records or inconsistent encoding of dates or names all have a negative impact on the use of your metadata but also on the implementation of linked data methodologies.”

“Reconciling” looks at controlled vocabularies – classification schemes, subject headings and thesauri. These controlled vocabularies allow “greater precision and recall during search and retrieval within an information system” and “within the context of linked data, they also allow connections to be created between collections.”

“Enriching” focuses on getting value out of non-structured metadata, making use of named-entity recognition (NER) for enriching existing metadata.

“Publishing” looks at best practices for sustainable publishing, including issuing URLs and managing them in the long term. The authors conclude that “URLs are the cornerstone of your linked data real estate.” They believe it is the core business of libraries, archives and museums to preserve and give access to collections, and so “maintaining solid URLs should therefore be a priority.”

The case studies come from a wide range of institutions: Schoenberg Database of Manuscripts, Powerhouse Museum, the British Library and Cooper-Hewitt National Design Museum. The case studies are well presented and easily identified in the text by means of a grey vertical bar in the margin.

Overall, I found this to be a challenging yet rewarding read and the work is a worthy contribution to the study of the key concepts of linked data and how they can be practically applied to existing metadata.

Neil Nicholson,

National Library of Scotland