HIDER, Philip. Information Resource Description: Creating and Managing Metadata. London: Facet, 2012.
220 pages. ISBN 978-185604-667-1.

"The nice thing about standards is that there are so many of them to choose from" (Andrew Tanenbaum). At first I expected this book to be an introduction to the various standards for metadata and how they are used. As such it would be of limited use to researchers. It does indeed cover a range of standards for an impressive range of media: library materials, archives, museum objects, etc. In fact I wonder whether there are any information-related courses or specialists of the breadth to cover all of these. Specialists will require more detail on the standards that relate to them, e.g. librarians on MARC formats and archivists on MAD, but this book is more concerned with thinking about the principles. It helps to understand how different standards apply to different levels, which is a frequent source of confusion. IFLA's FRBR (Functional Requirements for Bibliographic Records) model with the four levels of items, manifestations, expressions and works is also discussed early on and the "traditional" library cataloguing structure of ISO 2709 or XML for the overall record structure, MARC for the numbering of fields, AACR for the content of fields, and e.g. LCSH for the content of the 650 field are also covered. It has a bibliography of eight pages including an impressive range of standards at different levels, many with URLs, for further reading.

However, the book is much more than an introduction to the standards. It has a very thoughtful discussion of the value of metadata in relation to the cost of creating it (though with no suggestion of how to measure either). In the age of content-indexing services such as Google is there any place for human indexers/metadata creators? As Dawson and Hamilton say, "There are many reasons for the success and public acceptance of … Google, but metadata is not one of them." The Library of Congress may be at the edge of a slippery slope since it "commissioned a report which questioned the need for LCSH, in the light of undergraduate use of search engines such as Google".

Of course, one of the aims of adding metadata to web pages is to try and improve your rankings in search results listings – search engine optimisation (SEO). This is a rather different approach to metadata from merely describing the subject to facilitate retrieval and may be the context where web editors first become aware of "metadata". SEO is discussed briefly in one paragraph on page 105 but it is not considered important enough to appear in the index. Hider asserts that "meta tags ... do not always feature prominently in search engine algorithms" though search engine providers are understandably protective of their algorithms.

To some extent, clever search algorithms that look for singulars and plurals and misspellings, and offer the option of searching separate words or phrases, remove some of the need for controlled vocabulary in indexing. They rarely help with the homonym problem though: Wikipedia can ask for disambiguation of cells (in prisons), cells (in biology) and cells (electrical) but Google cannot!

Another new approach to retrieving information is social tagging. Post-modernism sees everyone's idea of what a resource is about as valid as anyone else's. Maybe there is no objective "subject" of a document that only a trained indexer can tease out but each "user" will consider it relevant to different needs. This is the philosophy behind social tagging where each contributor can add their two-pennyworth of what they think the document is relevant to – with no concession to vocabulary control. How can we apply the traditional measures of recall and precision when relevance itself is subjective?

Cataloguing, classification, indexing and perhaps abstracting have traditionally been seen as core activities of the librarian and information specialist. Cooperative cataloguing and the national MARC service have led to reduced need for and deskilling of cataloguers and classifiers as far as books are concerned. With information on the web, there is so much of it that content-based services seem to be the only option. But at the same time there are projects underway which aim to organise knowledge on a scale and with a level of detail hitherto inconceivable (except perhaps by Otlet and La Fontaine with their Répertoire Bibliographique Universel). If computers are to be able to produce factual answers, rather than links to places where answers may be found, they need data to be formatted in a highly structured and standardized way. This is the vision of the Semantic Web. Ontologies express relationships between concepts in a much more rigorous way than the "broader term/narrower term/related term" approach of traditional thesauri and classification systems. Geonames is a database with over eight million place names with exact co-ordinates, distinctions such as "populated place", "second-order administrative division", etc. The Virtual International Authority File aims to match and link name authority files on an international scale. DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia. Schema.org provides a collection of schemas, i.e., HTML tags, which webmasters can use to mark up their pages in ways recognized by major search providers. For example, for people there are properties defined such as "birthdate", "colleague", and "spouse". The extent to which existing metadata, metadata standards, and metadata specialists within the library and information world can contribute to these "linked data" projects and the development of the semantic web will surely provide much fruitful ground for research.

Andrew Buxton
Lewes