Harvey, Ross (2010) *Digital curation: a how-to-do-it manual.* London: Facet. xxii, 225p. ISBN 978-1-85604-733-3. Price £44.95 (£35.96 to CILIP members)

Ross Harvey is exceptionally well qualified to provide this timely guide to a field of strategic and operational importance to researchers, librarians, archivists and others in the scholarly information community. His recent positions as research fellow and visiting professor at centres of expertise in the UK and the US, following on from his academic career in Australia, have put him in the vanguard of thinking and practice in this area. As Harvey notes, digital curation is applicable to a wide range of digital objects, but it is developments in e-research or 'cyberscholarship' (as it is known in the US) and particularly the shift from laboratory-based experiments to computer-based data-driven research that have put the management – more specifically the *curation* – of digital datasets on the agenda of government agencies, research sponsors and scholarly publishers, as well as presenting opportunities and challenges for researchers and information professionals.

The Preface identifies the book's intended audience as anyone who creates, uses or curates data, but specifically librarians, archivists and students of these professions, as well as scientists and scholars involved in planning and conducting research projects. The author's stated intention is for the chapters to be read consecutively by those wanting an overview of the field or consulted selectively for general background and specific advice on the subject. The book has several features that add to its value as both an educational resource and a reference tool, including side-bars/boxes highlighting key points and tables comparing different frameworks (labelled as 'figures' in the text), in addition to clear headings and sub-headings that make it easy to find relevant sections. Each chapter displays its structure on the opening page by listing the main sections in a side-bar. There is also a four-page list of abbreviations at the front (particularly useful in this acronym-laden field) and a detailed eight-page index at the back. The whole work is extremely well supported by references: readers are continually directed to articles, briefing papers, case studies, guides, handbooks, reports and standards for additional information and illustration.

The book is arranged in three parts, made up of four, four and seven chapters respectively. The first part, 'Digital Curation: Scope and Incentives' sets the context for the rest of the text. Chapters 1 and 2 explain the concept of digital curation and the changes that have taken place in the academic research arena that have made digital curation a priority. Harvey promotes a broad interpretation of the concept, which goes beyond other definitions cited:

> 'Digital curation is concerned with actively managing data for as long as it continues to be of scholarly, scientific, research, administrative, and/or personal interest, with the aims of supporting reproducibility, reuse of, and adding value to that data, managing it from its point of creation until it is determined not to be useful, and ensuring its long-term accessibility, preservation, authenticity, and integrity' (p.8).

Chapter 1 also explains how curation is needed to ensure 'digital continuity', summarises its benefits and incentives, and outlines the roles and tasks involved, concluding with a usefully full summary of key points covered in the chapter. Chapter 2 describes the characteristics of new forms of research, citing examples from a range of disciplines, before considering the implications for information service organisations, including the skill sets and training needed, at one point suggesting the emergence of 'a new profession' (p. 25). Chapters 3 and 4 provide additional clarification of the scope and focus of the book, by introducing the conceptual model

that frames the chapters in Parts II and III – the Digital Curation Centre's (2008) *DCC curation lifecycle model* – and revisiting the definition of the term 'data' as used in the digital curation field, again stressing the need for a broad interpretation that includes all disciplines and accommodates continuing developments in policy and practice. The explanations of the different ways that data can be characterised and categorised in terms of their origin, status, function, subject, etc. are particularly valuable here, together with the discussion of the need to include among the digital objects considered for curation both the products resulting from the manipulation of data and the workflows (automated scientific processes) used to search, integrate and analyse data sets. Both chapters relate the frameworks adopted in the book to others found in the literature.

Each chapter in Part II, 'Key Requirements for Digital Curation', is devoted to one of the four Full Lifecycle Actions of the DCC model: *Curate and Preserve*; *Description and Representation Information*; *Preservation Planning*; and C*ommunity Watch and Participation*, though the chapter titles differ slightly from the wording used in the model. Chapter 5 clarifies the distinction between 'digital curation' and the activities within digital curation that constitute 'digital preservation' and then discusses the roles of different stakeholders in the curation process. Interestingly, librarians do not feature in this chapter, where the focus is more on the tasks of domain experts, typically scientists. Chapter 6 is the longest in the book at 24 pages and also has the most dense technical content, defining the different types of metadata needed (administrative, descriptive, technical, structural and preservation) and how they relate to and overlap with each other, then describing the key standards and specifications in the field, before moving on to the different classes and sub-classes of 'representation information', a concept less familiar to librarians and archivists. 'Preservation Planning and Policy' and 'Sharing Knowledge and Collaborating' are among the more straightforward chapters, but include useful brief discussion of the costs of curation (Chapter 7) and the intrinsically collaborative nature of digital curation (Chapter 8).

The seven chapters in Part III, 'The Digital Curation Lifecycle in Action', cover the eight Sequential Actions of the DCC model, with the final chapter covering both *Access, Use and Reuse* and *Transform.* These chapters together take up more than half of the book and form the core text for hands-on practitioners in the field and others venturing into this territory at a practical level. The chapter titles generally reflect the terminology of the model, but with amendment to more user-friendly language in a few cases. Thus *Conceptualise* becomes 'Designing Data' (Chapter 9) and *Appraise and Select* becomes 'Deciding What Data to Keep' (Chapter 11); surprisingly, perhaps, no plain English substitution is made for *Ingest.* The three Occasional Actions of the model (*Reappraise, Dispose* and *Migrate*) are discussed in Chapters 11 and 13 ('Preserving Data'), relating them to the most relevant Sequential Actions, which is all clearly explained in the introduction to Part III.

The chapters on appraisal/selection, preservation and storage are the most substantial here, but the treatment of all the Sequential Actions is thorough and clear. The importance of documented policies as the basis for sound procedures and the central role of standards and tools are recurring themes throughout Part III. Useful features include a checklist for data management in Chapter 9, sample questions for data interviews (Chapter 11), simplified and expanded versions of key steps in the ingest process (Chapter 12), a comparative table of preservation methods and a list of digital curation tools (Chapter 13), discussion of repository software options (Chapter 14) and diverse examples of annotation (Chapter 15).

Real-world examples from the scientific research and research library communities are used throughout the work to illustrate recommended practices, taken mainly from Australasia, the UK and US. The library examples are predominantly from national libraries, with relatively few representing university library experience, a notable exception being the set of data interview questions, drawn from Purdue University Libraries (Witt and Carlson, 2007). The book could have been enhanced by including more examples drawn from academic libraries. Readers from that sector will probably be surprised not to find more extensive discussion of university library practice, especially in view of the recent upsurge in published case studies, particularly from the US. The author was unlucky with his timing as the book evidently went to press a few months before publication of the latest report from the Association of Research Libraries, which contains six such case studies (Soehner et al., 2010). However, some of these cases had previously been reported in the literature and could have been referenced, including Johns Hopkins University (Choudhury, 2008) and MIT (Gabridge, 2009), in addition to fuller accounts of Purdue (Garritano and Carlson, 2009; Witt, 2008).

The book has a companion website[1] that provides a total of 25 checklists, 2 downloadable templates and more than 50 links to documents and websites for 12 of the 15 chapters. The section on 'Keeping Up-to-Date' in Chapter 8 is supported by 27 links, but provision of links for other chapters is highly selective with only a few items linked per chapter, in comparison with the extensive lists of references for many chapters in the book. Unfortunately, the usual problems of broken links and out-of-date/superseded material were already evident only a few months after publication, with links not working for two of the four source documents for checklists in Chapter 9 and the link for the DCC Data Management Plan Template on the Chapter 7 page taking users to the 2009 draft for consultation, rather than the 2010 post-consultation version. The website invites visitors to advise the author of other web-accessible checklists and templates, but it is not clear what the publisher's plans are for maintaining, updating or developing this site as a web resource, which could become an issue in a field that is continuously developing. It is obviously much quicker to be able to click on a web-page link to follow up topics of interest, but potentially frustrating for readers to find so few of the references in the book actually included on the web pages.

The publishers justifiably claim that this text is 'the first one-stop resource in digital curation' and as such it is sure to attract a large readership among its primary audience, given current interest in the topic around the world. The author provides an excellent step-by-step guide to this vital new area of practice, which should demystify the subject for those put off by the daunting array of technical vocabulary and huge number of acronyms encountered in specialist discussions. Although the book deals with technical issues in several places, it is written in straightforward non-technical language and does not assume a high level of technological knowledge. Indeed, one of the key messages of the text is that organisational, legal, political and social issues are crucially important to success in the digital curation field. There is some repetition of content between and within chapters, but this generally provides useful reinforcement and helps to make each chapter relatively self-contained, adding value to use as a reference resource. The book will undoubtedly be useful not just to students and practitioners at the sharp end of operations, but also to more senior staff with policy and strategy roles, as well as

---

[1] http://www.neal-schuman.com/curation/

academics and educators in the field and those with responsibility for workforce development. It is a valuable contribution and strongly recommended.

Sheila Corrall
Professor of Librarianship & Information Management
Information School, University of Sheffield, UK

References

Choudhury, G. Sayeed (2008) 'Case study in data curation at Johns Hopkins University', *Library Trends,* 57(2), Fall 2008, pp.211-220.

Digital Curation Centre (2008) *DCC curation lifecycle model.* Edinburgh: Digital Curation Centre, 2008. http://www.dcc.ac.uk/resources/curation-lifecycle-model

Gabridge, Tracy (2009) 'The last mile: liaison roles in curating science and engineering research data', *Research Library Issues,* 265, August 2009, pp.15-21.
http://www.arl.org/bm~doc/rli-265-gabridge.pdf

Garritano, Jeremy R. and Carlson, Jake R. (2009) **'**A subject librarian's guide to collaborating on e-science projects', *Issues in Science and Technology Librarianship,* 57, Spring 2009.
http://www.istl.org/09-spring/refereed2.html#15

Soehner, Catherine, Steeves, Catherine and Ward, Jennifer. (2010) *E-Science and Data Support Services: A Study of ARL Member Institutions.* Washington, DC: Association of Research Libraries, 2010. http://www.arl.org/bm~doc/escience_report2010.pdf

Witt, Michael (2008) 'Institutional repositories and research data curation in a distributed environment', *Library Trends,* 57(2), Fall 2008, pp.191-201

Witt, Michael and Carlson, Jake R. (2007) 'Conducting a data interview', *Poster presented at the 3rd International Digital Curation Conference, 11-13 November 2007, Washington DC.* Available at: http://docs.lib.purdue.edu/lib_research/81/.