# City Research Online

# City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

**Permanent repository link:**  https://openaccess.city.ac.uk/id/eprint/17311/

**Link to published version**: https://doi.org/10.1108/lm-01-2017-0009

# 'The story of data': a socio-technical approach to education for the data librarian role in the CityLIS library school at City, University of London

**Lyn Robinson and David Bawden**

## Abstract

### Purpose
This paper describes a new approach to education for library/information students in data literacy - the principles and practice of data collection, manipulation and management - as a part of the Masters programme in library and information science (CityLIS) at City, University of London.

### Design/methodology/approach
The course takes a socio-technical approach, integrating, and giving equal importance to, technical and social/ethical aspects. Topics covered include: the relation between data, information and documents; representation of digital data; network technologies; information architecture; metadata; data structuring; search engines, databases and specialised retrieval tools; text and data mining, web scraping; data cleaning, manipulation, analysis and visualization; coding; data metrics and analytics; artificial intelligence; data management and data curation; data literacy and data ethics; and constructing data narratives.

### Findings
The course, which was well-received by students in its first iteration, gives a basic grounding in data literacy, to be extended by further study, professional practice, and lifelong learning.

### Originality/value
This is one of the first accounts of an introductory course to equip all new entrants to the library/information professions with the understanding and skills to take on roles in data librarianship and data management.

**Introduction**

A role for librarians, and other information professionals, which is of considerable and increasing importance is the handling of data resources; on behalf of their users, and for their own purposes. This role, or perhaps it is better to say spectrum of roles, parallels that in the more traditional world of text and image resources. In supporting users, this ranges from a concern with the overall institutional, or even wider, policies for the management and curation of datasets of all kinds, to assisting an individual user with the detail of small-scale data handling and analysis. It also includes the collection, analysis, management, and use of data relating to library operations, and their use as metrics for service evaluation and improvement; an extension of the well-established 'library statistics'. The recent great expansion of the amount of available, and of public and institutional awareness of the importance of data, lends an urgency to the need for library/information specialists to be fully aware of the new 'data dimension' to their work, and this certainly amounts to a new role for librarians, in line with the theme of this Special Issue. As Ekstrøm et al. (2016) write "Imagine a librarian armed with the digital tools to automate literature reviews for any discipline, by reducing thousands of articles' ideas into memes and then applying network analysis to visualise trends in emerging lines of research. What if your research librarian could then dig deeper and use [a digital tool] to map in which sections of articles your key research terms appear? Imagine the results confirmed that your favourite research term almost never appears in the results sections, but cluster only around introductions and perspectives? And what if the librarian did not stop there, but zoomed into the cloud of data with savvy statistics, applying the latest text and data mining techniques to satisfy even the most scrutinising scientific mind, before formulating an innovative research question?"

Not all librarians, even in academic and research settings, will become data specialists to this extent, although many certainly will. But all library and information professionals, in all sectors, will need to gain at least a basic appreciation of the issues around data, both technical and socio-ethical.

This role certainly exists now, but will become of greater significance and ubiquity in future years. As Kirkwood (2016, p.275) puts it "Data are nothing without analysis, and many librarians currently lack the data fluency to work confidently in a world of dynamic content creation ... Librarians need both to re-skill and to change their self-identification and the philosophy that underlies it, if they are to achieve confident data fluency." This need for many, if not all, librarians to become more confident in dealing with data, a role which only a few years ago would be relevant to very few within the profession, is a vital one. The issue is not merely one of technical competence, important though that is, but of a confident appreciation of all the issues surrounding the good use of data, including the legal and ethical; much as librarians have traditionally had a confident appreciation of text-based publications.

If librarians - in general, and beyond a few specialists and enthusiasts - are to be effective in this new role, professional education will have to adapt accordingly; see, for example, the surveys of data-focused provision in courses in the US (Tang and Sae-Lim 2016) and in China (Si, Zhuang, Xing and Guo 2013). In the US, courses focusing on aspects of data science, data

handling and data management are offered within most educational programmes for library/information specialists, particularly, though not exclusively, in the iSchools.

One response is to provide programmes which specifically prepare students for the new data-centric roles, such as data librarian, data steward, data curator, research data manager and data archivist. Such programmes necessarily focus strongly on the development of technical and managerial skills of data handling, and are aimed at students who are aiming at a clearly data-focused career within the library/information sector. Examples of these are the programmes offered by the iSchools at the University of Pittsburgh (Lyon, Mattern, Acker and Langmead 2015), and at the University of Sheffield (University of Sheffield 2017).

Another response, which is the rationale for the course described in this article, is to adapt curricula to ensure that all new entrants to the library profession are given at least a basic foundational understanding of both the technology of data handling and management, and its social and ethical implications. The two aspects are of equal importance, and cannot sensibly be separated. Without a detailed and practical appreciation of the technical issues, consideration of social and ethical matters will necessarily be ungrounded and general; while without a socio-ethical appreciation it will be difficult for students to understand how technical skills should best be applied. For library/information professionals dealing with data in any respect, while technical competence is a necessity, it must be framed within an understanding of the social and ethical - and indeed the wider cultural and political - environment.

This paper describes an initiative, following the latter approach, within the library/information science Masters programme at City, University of London (CityLIS). This involves the repositioning of an introductory information technology (IT) course within the programme as a course dealing with data in all its aspects of relevance to the library/information professions, and from a socio-technical and ethical perspective.

**The data challenge for librarians**
Of the many changes and challenges impacting on the work of the library and information professional, the 'data deluge' is certainly among the most significant. The greatly increased amount and diversity of data available is one of the most important changes in the information landscape. This applies both to the very large and heterogeneous datasets which tend to termed 'Big Data', and to the smaller, but no less important, bodies of data collected for specific purposes (Sugimoto, Ekbia and Mattioli 2016; Borgman 2015).

The significance of data in the library/information context is two-fold.

First, information professionals may need to become involved in data support, research data management, data curation, data governance, data quality evaluation, data citation, data literacy training, and similar activities, as a part, or all, of their professional remit (Koltay 2015, 2016; Rice and Southall 2016). This may involve, at its most formal: assisting with, or managing, research data management policies and plans (Briney 2015); developing and managing data repositories; overseeing a data curation programme (Nielsen and Hjørland 2014; Oliver and Harvey 2016); designing training programmes for data literacy and associated skills, including basic coding, in environments including university, school

and public libraries (MacMillan 2015; Carlson, Nelson, Johnson and Koshoffer 2015; Crystle 2017); or dealing with data within an overall framework of digital scholarship (Borgman 2015; Mackenzie and Martin 2016). Or it may, in a less formal way, involve giving advice to individual users on how best to deal with their data, in the way that librarians have always advised on dealing with bibliographic references. Becoming, in part or in whole, a data librarian, in Rice and Southall's terminology, is simply a new extension of the information provision/information management function, albeit that it may a new role description.

Second, it is important for information professionals, even if they have no special role in helping their users deal with data, to be able to handle data of all kinds confidently for their own purposes; to use data analytics to improve their library services, for example (Farmer and Safer 2016; Kirkwood 2016; Showers 2015). When these two developments are considered together, it is clear that new entrants to the information professions must be equipped to deal as confidently with data, in its variety of forms, as they have traditionally dealt with text information. Achieving such data confidence means having a conceptual understanding of data, and the issues around it, plus the technical capabilities of 'data scraping' and 'data wrangling': the abilities to find, extract, collect, clean, organise, analyse, and present data.

Furthermore, there are two inter-related aspects to the kind of data fluency that the new environment demands of information professionals: the technical, and the social and ethical. There is little point in a librarian being able to code, to scrape data from websites, to clean and analyse datasets, and to produce metrics on demand, if they are unfamiliar with the legal requirements of, and ethical considerations implicit in, what they are doing. But equally, there is little point in such a person being able to fluently debate the social and ethical niceties, if they are unable to get their data they need, in the form they need it in, and to draw from it the meaningful information that it is of use. The two go hand in hand, and the understanding of data that the library/information professional must possess must be a socio-technical understanding, enabling them to deal with data with technical competence and with ethical confidence. There is, of course. also a legal dimension to the proper use of data; this is mentioned where necessary in the course described here, but a full treatment of legal issues comes in courses elsewhere in the City programme, dealing with information law.

The importance of these issues has been emphasised repeatedly, as may be shown by the following examples. The sheer volume of data to be dealt with is illustrated by the general acceptance that we have entered the 'zettabyte era', in which annual data traffic on global networks exceeds the zettabyte level (Cisco 2016, Floridi 2014). In response to this, the UK government has explicitly recognised the importance of data literacy as a way of helping non-data specialists make the most of data science (Parkes 2016), while the US National Information Standards Organization (NISO) is planning training webinars for 2017 putting data literacy on a par with digital literacy (NISO 2017).

In the library sector, a bibliography on research data curation noted 560 items published between 2009 and 2016 (Bailey 2016). 'Dealing with data' was named as one of '5 technical skills that information professionals should learn', according to an entry on the CILIP (Chartered Institute of Library and Information Professionals) blog in March 2016

(Pennington 2016). This emphasised the need to deal with four distinct types of data: structured (e.g. spreadsheets and relational databases); semi-structured (e.g. files of metadata records); unstructured (without any table or field structure and encompassing big data); and linked data. Similarly, 'Using social media analytics' was named as one of the 'top five library technology topics' by the Techsoup for libraries' blog in December 2016 (Gilbert-Knight 2016).

Training for librarians has begun to develop to match these perceived needs. To give three examples: the library of North Carolina State University hosts a week-long 'Data science and Visualization Institute for Librarians' (North Caroline State University 2017); the Library of Congress held a conference on 'Collections as Data' in October 2016, with the two main themes that digital collections are composed of data that can be acquired, processed and displayed in many ways, and that we should always remember that data is derived from, and manipulated by, people (Ashenfelder 2016); and the American Library Association and Google, though their *Libraries Ready to Code* project, are seeking to equip librarians to teach coding and data handling in public and school libraries (American Library Association 2017).

These kinds of developments support the need for all librarians to have a solid socio-technical grounding in data issues.

**IT teaching at CityLIS**
An introductory information technology course has been offered as a compulsory part of the library/information programmes at City since Masters level teaching in the subject was established in its current structure in the late 1980s (Robinson and Bawden 2010). This course has always been seen as an introduction to basic concepts, and a preparation for more specialist courses. [Note that in this paper we use the term 'programme' for the whole Masters scheme of study, and 'course' for this specific part.]

This course was initially called 'Computers and communications technology', and the very broad syllabus was:

> Information systems and technology. An introduction to computers, hardware, software, operating systems, programming languages, software packages, databases, word processing, spreadsheets. Terminology and basic concepts of telecommunications. Telecommunications-based systems, including telex, fax, electronic mail, teleconferencing, videotex, electronic journals, document delivery systems, office automation. Hard copy techniques, including copying, duplicating, printing, graphic design and composition, desktop publishing. Microforms and their applications. Introduction to systems analysis.

In 1996, the Masters programme was restructured on a modular basis, and the course renamed 'Information technology', with a greater digital emphasis. By 2003-04, the course was named 'Data Representation and Management', and by then focused entirely on digital systems. Its emphasis was on software systems for handling various kind of information: text handling and word processing systems, spreadsheets, web authoring, databases, etc. In 2008, the course was renamed 'Data and Information Technology and Architecture' and

shortly afterwards 'Digital Information Technologies and Architecture'; de-emphasing data handling and taking a wider perspective. Its aim was to "provide the technical background required to store, structure, manage and share information effectively". It still included material on specific kinds of software, but was increasingly focused on web-based systems, search engines, blogs and wikis, semantic web, information retrieval, etc., and on information architecture, and issues such as open access and repositories.

In academic year 2016-17, this course was given a major overhaul. It was realised that the introductory material on software use was no longer necessary, while the detailed material on web-based systems, retrieval and information architecture was better left to later specialist and elective courses. Eliminating this material allowed for a new focus on the handling of data in all its aspects, as the essential background preparation for the new data roles mentioned above; a return to the data focus of earlier years, but with a very different treatment appropriate to the new environment. It was also felt essential to introduce a strong flavour of ethics, and social implications, hitherto missing in what was very much a technical course. The revised course, with its strongly socio-technical perspective, was renamed as 'Digital Information Technologies and Applications', to indicate that information architecture was not not such a central point. It took the strapline 'The Story of Data', to match another part of the programme called 'The Story of Documents'.

**The story of data**
The stated aim of the restructured course is to "provide the technical and philosophical background required to collect, store, describe, structure, manage and share information effectively in the digital society", by engaging with the deluge of digital data, and distilling information from it. The theme "Finding the I in data" is emphasised, with a double meaning: finding meaningful information (I) in data, and also considering how data represents or misrepresents us as individuals (I). There is also a strong focus on implications for library/information applications and issues, to ensure that the course does not become a generic 'data science lite'. In drawing up the syllabus, we were particularly influenced by North Carolina's 'Data Science and Visualization Institute for Librarians' mentioned above, and by modules in the Oxford Internet Institute's Masters programme in 'Social Science of the Internet' (Oxford Internet Institute 2017). We drew from these programmes ideas for both the balance of technical and conceptual material, and the balance of practical activities with consideration of conceptual and managerial aspects, as well as the general 'flow' of the course. More specifically, they influenced our decisions to use the Python language to illustrate the value of coding, and to use examples of scraping data from the Web whenever possible.

Although there is no single recommended text for the course - the material is too broad and diverse - the technical content is roughly matched by Herzog (2015) and the socio-ethical content by Floridi (2014).

For the central concept - data itself - we follow Floridi's definition: data is any discernible difference, or lack of uniformity; information is well-formed, meaningful and truthful data (Floridi 2010).

The course is organised in ten sections: their titles are stated here to show the trajectory of the story, and discussed below:

The story of data
1       Finding the 'I' in data
2       You will be assimilated
3       Data about data
4       Taming of the data
5       Searching for the data
6       Working with the data
7       Counting the data
8       The meaning in the data
9       AI: the data will replace you
10      Making data work

Each section includes two class sessions - presentations, demonstrations and practical work - plus significant independent student work; the whole course (a 15 UK credit module) accounting for a nominal 150 hours student work. This is sufficient to ensure that all students have the opportunity to gain an appreciation of each topic, conceptually and practically, and to be in a position to learn more, either during their studies or in the workplace. For some sections, guest lecturers from institutions such as the UK Digital Curation Centre, Altmetric, and CILIP offer the viewpoint from the world of practice.

Considering each section in turn, we now briefly outline its content.

### 1      Finding the 'I' in data

This introductory section considers the modern phenomenon of the data deluge, and its implications for the individual. It considers: the relation between data, information and documents (Floridi 2010); the historical development of computer systems, and the ways in which computers represent and handle data - Turing and von Neumann architectures, bits and bytes, and coding systems (Ince 2011); and socio-technical issues, particularly for the library/information profession. This section establishes the conceptual framework for the course, and provides the understanding of basic issues needed by any librarian dealing with data.

### 2      You will be assimilated

This section introduces networks and digital network technologies, specifically the internet and the web, and the standards and protocols which underlie them, most notably TCP/IP and HTML. The concepts of the web and web pages are used to introduce some basic ideas of information architecture (Rosenfeld, Morville and Arango 2015). Some social and ethical implications of data transfer and sharing - including individual presence and privacy online, digital divide, net neutrality, and the implications of the design of network infrastructures - are considered. This establishes an understanding of the web environment in which virtually data in the library context resides.

### 3   Data about data

This section considers the ways in data forms documents (Furner 2016), and how different kinds of documents are defined, described and organised, leading to an introduction to metadata standards and applications. Following the approach of Pomerantz (2015), this treats metadata very broadly, giving some attention to bibliographic and web resource metadata, but focusing equally on metadata for datasets. This provides a link between the metadata concepts familiar to librarians to their application in the less-familiar dataset context.

### 4   Taming of the data

This section considers the structuring of data into organised data files of various kinds: flat files, CSV files, database structures including relational, and standards, including XML, RDF and linked data. This leads to a discussion of the processes of data management, for research and for other purposes, and of data curation (Briney 2015; Oliver and Harvey 2016). A conceptual understanding of, and an ability to work with, data files of these kinds is fundamental to the success of librarians in confidently dealing with data collections.

### 5   Searching for the data

This section considers how to find data of various forms, building on early discussions of data structure. It covers the range of search tools for various forms of data collection: search engines, relational database systems and SQL, full text bibliographic search systems, and other specialised retrieval tools. Carlson, Nelson, Johnson and Koshoffer 2015). It subsumes the text retrieval and bibliographical retrieval systems familiar to most librarians within the broader framework of systems with retrieve data of all kinds.

### 6   Working with the data

This section focus on the ways data can be collected from web services and APIs, such as Twitter, and then cleaned, manipulated and analysed; what is sometimes termed 'data scraping' and 'data wrangling'. Software such as Hawksey's Tagsexplorer and OpenRefine (Groves 2016) are used to illustrate collection, summarisation and visualisation. A facility with this kind of process will be particularly valuable to librarians seeking to become experts in helping their users deal with data issues, as it is becoming a wide-spread form of data usage.

### 7   Counting the data

This section examines data metrics, introducing basic analytics, basic bibliometrics (as an introduction to the study of bibliometrics laws and applications later in the programme), and altmetrics (Tattersall 2016). While counting data is now technically quite straightforward, we ask what are we measuring when we measure data, and what does it mean? Again, this is an extension of issues familiar to librarians - the bibliometrics of conventional publication - into the less-familiar data realm.

### 8   The meaning in the data

This section examines tools for exploring data to find meaning in it, including tools for text and data mining, and for visualization. Standard packages - Wordle, Tagxedo and Voyant Tools (Megan 2014, Moorfield-Lang 2010) are used for collection and analysis of both structured and unstructured data from the web. There is a basic introduction to coding in

the Python language, including use of general and specialised subroutine libraries, web scraping via API wrapper, and regular expressions. The aim is to illustrate the purpose of coding, and where it offers advantages over the standard packages, with examples of library/information applications. The ability to undertake basic coding is now a valuable skill in many library contexts, including modifying bibliographic records, enriching metadata, converting record formats, customising interfaces, and linking systems. This section also considers the discipline of digital humanities, which has provided many of these tools, and its relationship to LIS (Svensson and Goldberg 2015; Robinson 2016).

### *9 AI: the data will replace you*

This section examines artificial intelligence (AI), from popular visions and historical developments to current practice, and implications for the information professions. Topics include machine learning, automatic indexing, tagging, classification and categorisation, artificial agents, web bots in general and chatbots in particular, and robots. Issues include whether librarians will really be replaced by robots, what the likely balance of the human and the digital will be, and what are some of the ethical implications, following the approaches of Boden (2016), and Floridi (2016, 2017). Some understanding of these issues is essential for new entrants to the library profession, as the impact of AI to all sectors will be significant.

### *10 Making data work*

This section, in a sense, circles back to the first section, considering the importance of data handling and management for the future library/information professional. How can they best contribute to managing the data deluge, and how can data be used to improve, justify and show the impact of, library/information services? No attempt is made to give definitive answers to these questions; rather this section opens a discussion, to be continued throughout the CityLIS Masters programme.

All aspects of the learning context have been changed to emphasise the integration of the technical and social/ethical treatment of data issues. Previously the course had been run by formal lectures followed by practical classes in a computer room. The computer room classes have been abandoned in favour of using seminar room for all sessions, and encouraging students to bring and use their devices (laptops, tablets, smartphones) for short practical in-class exercises, which can be naturally integrated with presentation, and which encourages discussion and peer support. Practical exercises have been adjusted so as to be doable without any special hardware or software, by using standard web-based systems: Voyant Tools, Wordle, Tagxedo, Tagsexplorer, Openrefine, etc.  For the introduction to coding, which uses the Python language, we are able to recommend a choice of online tutorials for practice, including one which requires only a web browser, rather than any special software. Those students with a strong interest can, of course, take things further by using special purpose hardware and software available at the university.

The purpose of including the coding component is not to train the class to become efficient coders: that would be neither desirable nor feasible in the time available. It is not necessary that all librarians be coders, but it is necessary that they understand the nature and purpose of coding, and when and why writing code may be preferable to using prepacked software. In order to do this, it is necessary to have some practical experience of coding. This course

provides this, in the context of data collecting and processing, for those students who have not encountered coding before. For those who have, it provides an introduction to a language, Python, with a rich provision of libraries and subroutines for accessing and manipulating data of the kinds of most interest to library/information practitioners. The aim is not to try to develop professional programming skills, but to show coding as a tool for creative exploration of data, following the approach espoused by Montford (2016). Students needed a more in-depth treatment of programming can find it elsewhere in the programme, especially by following technology-oriented electives, and by participating in 'out of hours' option technology training. An example of the latter is CityLIS's hosting of the first Library Carpentry software training (Playforth 2015).

Background reading and resources for each section are designed to cover three perspectives: the technical; the socio-ethical, and the professional, outlining the implications for library/information professionals. While the sections are distinguished mainly by their technical content, the social and ethical concerns tend to overlap, since their principles are applicable in many aspects of information and data management (Floridi 2013, Floridi and Taddeo 2016).

The assessment for the course is an essay or report on a topic chosen by the student, but which must incorporate both technical and socio-ethical aspects. Students are also required to set up a blog, if they do not already have one, and use it to reflect on their learning as the course progresses, and also encouraged to use other forms of social media such as Twitter, so as to ensure that all are comfortable with communicating via digital media.

**Conclusions**

At the time of writing, the course had been given for the first time. Reaction from students, and from the expert practitioners acting as guest lecturers, suggests that this is an engaging and effective way of introducing students to the role of library/information professionals in managing data, understanding both the technology and its social and ethical dimensions. A more through and formal evaluation at the end of the academic year will influence the future direction of the course. The fact that is it is compulsory for all library/information students, and indeed is the first course they encounter in their studies, helps emphasise the importance of understanding data and its implications in all library/information contexts.

Data issues are clearly here to stay as a significant aspect of the work of all librarians, and other information professionals, and all entrants to the profession need a good socio-technical grounding as a basis for professional practice, and - vitally - continuing learning throughout professional life. We hope that this new CityLIS offering, which will be further developed over future years, will serve this purpose for our students, and may be a useful example to others.

## References

American Library Association (2017), Equipping librarians to code: ALA, Google launch ready to code university pilot programme, [blog post], available at http://www.ala.org/news/press-releases/2017/01/equipping-librarians-code-ala-google-launch-ready-code-university-pilot, accessed 20 January 2017.

Ashenfelder, M. (2016), Data and humanism shape Library of Congress conference, [blog post], available at http://blogs.loc.gov/thesignal/2016/10/data-and-humanism-shape-library-of-congress-conference/?loclr=eadpb, accessed 17 January 2017.

Bailey, C.W. (2016), *Research data curation bibliography (version 6)*, Houston TX: Digital Scholarship, available at http://digital-scholarship.org/rdcb/rdcb.htm, accessed 16 January 2017.

Boden, M.A. (2016), *AI: its nature and future*, Oxford: Oxford University Press.

Borgman, C.L. (2015), *Big Data, Little Data, No Data: Scholarship in the Networked World*, Cambridge MA: MIT Press.

Briney, K. (2015) *Data management for researchers: organize, maintain, and share your data*, Exeter: Pelagic Publishing.

Carlson, J., Nelson, M.S., Johnson, L.R. and Koshoffer, A. (2015), Developing data literacy programs: working with faculty, graduate students and undergraduates, *Bulletin of the Association for Information Science and Technology*, 41(6), 14-17.

Cisco (2016), The Zettabyte era - Trends and Analysis, [online], available at http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html, accessed 20 January 2017.

Crystle, M. (2017), Libraries and facilitators of Coding for All, *Knowledge Quest*, 45(3), 46-53.

Ekstrøm, J., Elbaek, M., Erdmann, C. and Grogorov, I. (2016), The research librarian of the future: data scientist and co-investigator, *LSE Impact of Social Sciences blog*, December 14 2016, available at http://blogs.lse.ac.uk/impactofsocialsciences/2016/12/14/the-research-librarian-of-the-future-data-scientist-and-co-investigator/, accessed 26 March 2017

Farmer, L.S.J. and Safer, A.M. (2016), *Library Improvement through data analytics*, London: Facet Publishing.

Floridi, L. (2017), Charting our AI future, *Project Syndicate*, [online], available at https://www.project-syndicate.org/commentary/human-implications-of-artificial-intelligence-by-luciano-floridi-2017-01, accessed 17 January 2017.

Floridi, L. (2016), Should we be afraid of AI?, *Aeon Essays*, [online], available at https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible, accessed 17 January 2017.

Floridi, L. (2014), *The fourth revolution: how the infosphere is reshaping human reality*, Oxford: Oxford University Press.

Floridi, L. (2013), *The ethics of information*, Oxford: Oxford University Press.

Floridi, L. (2010), *Information: a very short introduction*, Oxford: Oxford University Press.

Floridi, L. and Taddeo, M. (2016), What is data ethics?, *Philosophical Transactions of the Royal Society A*, 374: 20160360, http://dx.doi.org/10.1098/rsta.2016.0360

Furner, J. (2016), "Data": The data, in Kelly, M. and Bielby, J. (eds), *Information cultures in the digital age*, Wiesbaden: Springer VS, pp 287-306.

Gilbert-Knight, A. (2016), Your top 5 library technology topics, Techsoup for libraries blog (9 December 2016), available at http://techsoupforlibraries.org/blog/your-top-5-library-technology-topics, accessed 16 January 2016.

Groves, A. (2016), Beyond Excel: how to start cleaning data with OpenRefine, *Multimedia Information and Technology*, 42(2), 18-22.

Herzog, D. (2015), *Data literacy: a user's guide*, London: Sage.

Ince, D. (2011), *The computer: a very short introduction*, Oxford: Oxford University Press.

Kirkwood, R.J. (2016), Collection development or data-driven content curation? *Library Management*, 37(4/5(, 275-284.

Koltay, T. (2015), Data literacy: in search of a name and identity, *Journal of Documentation*, 71(2), 401-415.

Koltay, T. (2016), Data governance, data literacy and the management of data quality, *IFLA Journal*, 42(4), 303-312.

Lyon, L., Mattern,E., Acker, A. and Langmead, A. (2015), Applying translational principles to data science curriculum development, in *iPres 2015*, November 206 2015, Chapel Hill, North Carolina, available at http://d-scholarship.pitt.edu/27159/, accessed 17 January 2017.

MacMillan, D. (2015), Developing data literacy competencies to enhance faculty collaborations, *Liber Quarterly*, 24(3), 140-160.

Mackenzie, A. and Martin, L. (eds.) (2016), *Developing digital scholarship: emerging practices in academic libraries*, London: Facet Publishing.

Megan, W.E. (2014), Review of Voyant Tools, *Collaborative Librarianship*, 6(2), 96-97.

Montford, N. (2016), *Exploratory programming for the arts and humanities*, Cambridge MA: MIT Press.

Moorfield-Lang, H. (2010), Infographics: information gets visual, *Information Searcher*, 19(3), 15-16.

Nielsen, H.J. and Hjørland, B. (2014), "Curating research data: the potential roles of libraries and information professionals", *Journal of Documentation*, 70(2), 221–240.

NISO (2017) NISO two-part webinar: Digital and data literacy, available at http://www.niso.org/news/events/2017/webinars/sept13_webinar, accessed 20 January 2017.

North Carolina State University (2017), Data science and Visualization Institute for Librarians, [online], available at  https://www.lib.ncsu.edu/datavizinstitute, accessed 20 January 2017.

Oliver, G. and Harvey, R. (2016), *Digital Curation*, London: Facet Publishing.

Oxford Internet Institute (2017), MSc Social Science of the Internet, [online], available at https://www.oii.ox.ac.uk/study/msc, accessed 20 January 2017.

Parkes, E. (2016), Data literacy: helping non-data specialists make the most of data science. Government Digital Service blog post, available at https://gds.blog.gov.uk/2016/04/27/data-literacy-helping-non-data-specialists-make-the-most-of-data-science, accessed 20 January 2017.

Pennington, D. (2016), 5 technical skills information professionals should learn. CILIP blog (22 March 2016), available at http://www.cilip.org.uk/blog/5-technical-skills-information-professionals-should-learn, accessed 16 January 2016.

Playforth, C. (2015), Why the information profession needs Library Carpentry [blog post], available at https://blogs.city.ac.uk/citylis/2015/12/07/why-information-profession-needs-library-carpentry, accessed 20 January 2017.

Pomerantz, J. (2015), *Metadata*, Cambridge MA: MIT Press.

Rice, R. and Southall, J. (2016), *The Data Librarian's Handbook*, London: Facet Publishing.

Robinson, L. (2016), Are the digital humanities and library and information science the same thing? [blog post], available at https://thelynxiblog.com/2015/06/29/are-the-digital-humanities-and-library-information-science-the-same-thing/, accessed 17 January 2017.

Robinson, L. and Bawden, D. (2010), Information (and library) science at City University London: 50 years on educational development, *Journal of Information Science*, 36(5), 631-654.

Rosenfeld, L., Morville, P. and Arango, J. (2015), *Information architecture for the web and beyond (4th edn.)*, Sebastopol CA: O'Reilly Media.

Showers, B. (2015), *Library Analytics and Metrics*, London: Facet Publishing.

Si, L., Zhuang, X., Xing, W. and Guo, W. (2013), The cultivation of scientific data specialists: Development of LIS education oriented to e-science service requirements, *Library Hi Tech*, 31(4), 700–724.

Sugimoto, C.R., Ekbia, H.R. and Mattiolli, M. (2016), *Big data is not a monolith*, Cambridge MA: MIT Press.

Svensson, P. and Goldberg, D.T. (2015), *Between humanities and the digital*, Cambridge MA: MIT Press.

Tang, R. and Sae-Lim, W. (2016), Data Science Programs in U.S. Higher Education: An Exploratory Content Analysis of Program Description, Curriculum Structure, and Course Focus, *Education for Information*, 32(3), 269-290

Tattersall, A. (ed.) (2016) *Altmetrics: a practical guide for librarians, researchers and academics*, London: Facet Publishing.

University of Sheffield (2017), MSc Data Science, [online], available at http://www.shef.ac.uk/is/pgt/courses/ds#tab01, accessed 20 January 2017.