IJDC | *General Article*

Bridging the Data Talent Gap: Positioning the iSchool as an Agent for Change

Liz Lyon
School of Information Sciences
University of Pittsburgh

Aaron Brenner
University Library System
University of Pittsburgh

Abstract

This paper examines the role, functions and value of the "iSchool" as an agent of change in the data informatics and data curation arena. A brief background to the iSchool movement is given followed by a brief review of the data decade, which highlights key data trends from the iSchool perspective: open data and open science, big data and disciplinary data diversity. The growing emphasis on the shortage of data talent is noted and a family of data science roles identified. The paper moves on to describe three primary functions of iSchools: education, research intelligence and professional practice, which form the foundations of a new Capability Ramp Model. The model is illustrated by mini-case studies from the School of Information Sciences, University of Pittsburgh: the immersive (laboratory-based) component of two new Research Data Management and Research Data Infrastructures graduate courses, a new practice partnership with the University Library System centred on RDM, and the mapping of disciplinary data practice using the Community Capability Model Profile Tool. The paper closes with a look to the future and, based on the assertion that data is mission-critical for iSchools, some steps are proposed for the next data decade: moving data education programs into the mainstream core curriculum, adopting a translational data science perspective and strengthening engagement with the Research Data Alliance.

Introduction

This paper examines the role, functions and value of the "iSchool" as an influential and effective agent of change in the data informatics and data curation arena. The iSchool movement originated in 1988 when the three Deans of the information schools at the University of Pittsburgh, Drexel University and Syracuse University, joined together to form a group that was extended to include the Deans from the University of Michigan and the University of Washington. The group continued to expand from 2003 onwards; the Deans were united by their common view of the breadth of scope of "Information Sciences" and how best to communicate this view to the wider academic community, recognising that there were intersects with areas such as computer science and telecommunications, and a strong "trans-disciplinary" dimension (Larsen, 2010). The discussions focussed on identity issues and the term "iSchool" emerged as an effective descriptor for the work of this group, which currently features fifty-five institutions covering a geographic spread from the origins in North America to Europe, Asia and Australasia[1]. Today, the iSchools comprise a mix of information-centric academic departments and Schools, including Library and Information Science, Information Systems, Computer Science etc. The iSchool members and their programs collectively encompass many facets of information science, including information systems, knowledge management, telecommunications, human-computer interaction, librarianship, archival studies, cultural heritage, media and journalism. More recently, iSchools have embraced the digital curation agenda, integrating various specialisations within their curricula (Corrall, Kennan and Afzal, 2013).

We begin this paper by briefly looking back at the significant data curation trends and developments of the last ten years or so (starting in 2003), from the iSchool perspective. The second section will articulate a Capability Ramp Model for iSchools, based on three primary iSchool functions: education, research intelligence and professional practice. The Capability Ramp Model is illustrated in the next section by specific data-centric mini case studies from the University of Pittsburgh, but also draws on evidence and exemplars from other iSchools in North America, Europe and Australia. We close by looking forward and exploring future engagement, potential opportunities and collective impact of iSchools in the next (data) decade.

Reviewing the Data Decade:
The Emerging Data Talent Gap

In the last decade, data curation, data preservation and data science have emerged as major development areas cutting across all business sectors with significant impacts on education, industry and governments. The imminent "data deluge" was highlighted in the seminal paper by Hey and Trefethen (2003) and the associated data-intensive science mode proposed as the *Fourth Paradigm* by the late Jim Gray (2009). Whilst a more complete description and analysis of the data decadal timeline is beyond the scope of this paper, here we briefly highlight three key data trends which have particular resonance for the iSchool community: open data and open science, big data and disciplinary data diversity.

---

[1] iSchools: http://ischools.org

Many reports and papers have provided definitions, insight, analysis and recommendations on the changing practice, perceived value and challenges of open science to the varied data stakeholders who are actors in the scholarly communication and data publication process, and who are suppliers of the requisite data infrastructure and services (Lyon, 2009; Royal Society, 2012; Corrall and Pinfield, 2014). In order to realise the aspirations and full potential of the new publication modes (ideally with an upper case P as proposed by Callaghan, 2012), data need to be effectively collected, cleaned, documented, stored, identified, released and preserved for the long-term, as first class outputs or products of research. iSchools have established roles in promoting open scholarly communication modes; they contribute to the design, development, testing, evaluation and dissemination of innovative methodologies, tools and services that are components of a more robust and trustworthy global information infrastructure. Some iSchools, including the University of Pittsburgh, also focus on information assurance, cybersecurity and privacy issues.

The concept of big data as defined by Gartner in terms of the 3Vs: "Big data is high volume, high velocity and/or high variety information assets that require new forms of processing to enable enhanced decision-making, insight discovery and process optimization" (Laney, 2012), has also generated much commentary. The McKinsey Global Institute Report (Manyika et al., 2011) included analysis of the shortage of data analytics skills and stated that "[t]he United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings." More recently the UK government has identified "a shortage of skilled workers" in its *Seizing the Data Opportunity* white paper (Department for Business, Innovation and Skills, 2013), and the *Model Workers* report from Nesta (Bakhshi, Mateos-Garcia, and Whitby, 2014) highlighted "a severe shortage of UK data talent." Each of these latter reports also identified a role for higher education institutions in helping to develop critical data skills. Whilst acknowledging that these reports focussed primarily on data science skills aligned with data analytics, statistical analysis and computational modelling, there are many other data roles which can be associated with this trend. The term "data scientist" was used in the US National Science Board Report on *Long-Lived Data Collections* (2005) to describe "the information and computer scientists, database and software engineers and programmers, disciplinary experts, curators and expert annotators, librarians, archivists, and others, who are crucial to the successful management of a digital data collection." Lyon and Takeda (2012) have proposed a further interpretation of the term "data scientist" to include the variants which are articulated in Table 1 below (note that this is not an exhaustive list, the descriptions are indicative rather than definitive and there may be overlap in categories). Data scientists can follow varied career routes e.g. disciplinary research scientists augmenting their skills through an iSchool Program. They may be based within research teams or in the corporate sector.

An analysis of emerging job trends was presented by Larsen et al. (2014) and highlighted the growth in new positions which include some element of "data". Whilst some of the job postings are re-titled from earlier instantiations, such as business intelligence roles, others are completely new designations and reflect the recognition that there is a requirement to build capacity and capability in a wide range of data areas.

The final trend to be highlighted is a growing recognition of the disciplinary diversity of data and related workflows, processes and cultures embedded in data practice. Comparing the relatively well-established, well-funded and co-ordinated infrastructure and common practices in astronomy, with a much more distributed, artisanal and cottage industry approach observed in network sensing initiatives

Table 1. Family of data scientist roles.

Role	Focus	Typical Location
Data analyst	Business/scientific analytics, mathematics, statistics, modelling	Corporate Sector
Data archivist	Long term preservation, repository management	National Archive
Data engineer	Software development, coding, programming, tools	IT Company
Data journalist	Telling stories and providing news using visualisations	Newspaper Publisher
Data librarian	Advocacy, research data management, training	University or Research Institute
Data steward /curator	Curation, cleansing, annotation, selection and appraisal	Data Center

(Wallis, 2013; Borgman et al., 2014), illustrates the contrast between those communities at the head and those in the very long tail (Heidhorn, 2008). Understanding the varied data practices amongst the wide range of disciplines and sub-disciplines is now beginning to be revealed through third party observational studies, surveys and fieldwork.

iSchools and Data:
Developing the Capability Ramp Model

iSchools are uniquely positioned to be engaged and informed co-partners in many data initiatives through their three primary functions: education, research intelligence and professional practice:

- Education – iSchool faculty are responsible for the design, development and delivery of innovative data programs, courses and certificates to undergraduate, Masters and Doctoral students. This curriculum development in data programs is a relatively new initiative: Corrall et al. (2013) estimated that around 33% US iSchools were offering one or more curation course.

- Research intelligence – iSchool faculty carry out original research which may include the collection, analysis and dissemination of qualitative or quantitative information, software development and testing, field work and legal, ethical and ethnographic studies. An increasing number of studies are exploring aspects of data curation and research data management (RDM) in collaboration with domain scientists (Borgman et al., 2014).

- Professional practice – iSchool faculty may have had prior roles as practitioners and bring their experience of service-oriented positions to their scholarship. An in-depth understanding of current trends, operational practice and service issues is vital to ensure the relevance and currency of iSchool education programs.

iSchools also engage on a day-to-day basis with key data stakeholders:

- Student community – some of whom are aspirant data scientists who enrol on data programs, courses and certificates, seeking to gain the necessary skills and knowledge to equip them to secure and adopt the roles described in Table 1.

- Disciplinary faculty and domain experts – working in diverse research environments (universities, institutes, clinical settings, pharmas, industry).

- Professional data scientists – located in a range of organisations, such as libraries, archives, data centers, IT services, large multi-national corporates, small start ups etc.

The relationships between these entities are varied and complex. iSchool faculty maintain their engagement with libraries, archives and expert centers through partner programs, intern placements, joint projects, advisory boards and collaborative initiatives. They collaborate with faculty in other departments and schools in inter-disciplinary research grant proposals, projects, patents and company start ups, and interact on a daily basis with the student body. This is a privileged and unique position, and one that enables an iSchool to act as a highly effective agent of change in emerging areas, with a receptive body of student learners, and data-savvy research teams. In the current marketplace, the ability to develop new data capability and talent is of critical value to employers. These integral elements are brought together in a new functional model, which is supported by illustrative case studies from the iSchool at Pittsburgh, with additional exemplars from elsewhere.

The new model introduced here is based on the concept of "ramps", which was described by Atkinson et al. (2010) for the eScience environment as "the method to scaling interactions by reaching deeper into communities and reaching out to new communities." Two types of ramp were presented for data-intensive science: a Technological Ramp, such as the Rapid portal software, and an Intellectual Ramp exemplified by the myExperiment social website for sharing workflows. A third type of ramp is proposed to leverage the distinctive iSchool data environment: a Capability Ramp that provides a means by which an increase in (data) skills, capacity and practical experience can be achieved. The relationships between these entities are illustrated in Figure 1.
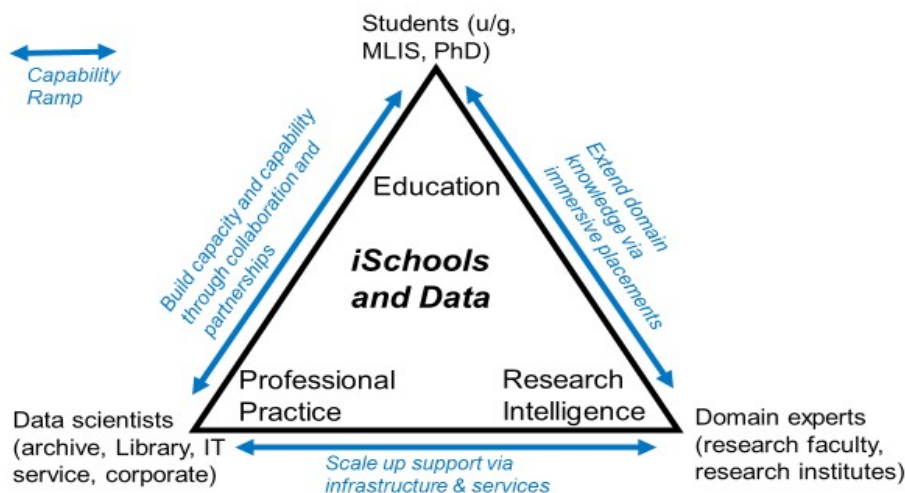

Figure 1. iSchool Capability Ramp Model.

iSchool Data Exemplars: Preliminary Evidence

Three mini case studies from the School of Information Sciences University of Pittsburgh (iSchool) are presented, which describe new data initiatives in each of the three functional zones (education, research intelligence and professional practice), and which demonstrate the potential of Capability Ramps.

The Immersive Experience: Addressing the Domain Disconnect

One of the biggest challenges for libraries and information services in providing research data management support is the diversity in disciplinary research practices, community standards and cultural norms. Generic information skills and digital competencies may not be sufficiently domain-focussed or at the depth of subject knowledge and expertise required to support data-intensive research. Similarly, when students bring prior knowledge from one particular subject area e.g. history, to bear on a very different data domain e.g. chemistry, we see a "domain disconnect" which may act as a barrier to capacity-building efforts. One possible strategy for addressing this issue is to implement internships in data centers (Palmer et al., 2014); an alternative strategy which is described in this paper, is to collaborate with local institutional research laboratories to offer an immersive experience to iSchool students.

At Pittsburgh iSchool, two new graduate data course specialisations have been developed with an immersive component. This approach builds on the initial immersive informatics pilot study, where a novel RDM training course was developed and delivered in a partnership between UKOLN Informatics, University of Bath and the University of Melbourne (Shadbolt et al., 2014). The first combined Masters (MLIS)/Doctoral course on Research Data Management covers a data timeline, the landscape of external policy drivers, data requirements, data management plans (DMPs), disciplinary data exemplars, data centers, advocacy and training, sustainability and costs, and legal and ethical issues. The second course on Research Data Infrastructures includes sessions on data storage, data sharing, data publication and citation, data discovery, data standards, data repositories, data preservation, citizen data, data science and further disciplinary data exemplars.

The immersive unit features the iSchool students going into a laboratory and working in pairs alongside a research scientist. The Department of Public Health and the Department of Epidemiology within the School of Medicine, and the RFID Center within the Swanson School of Engineering (all at the University of Pittsburgh), have each hosted students this year. The students and researchers are both briefed in advance with an interview outline and topics to cover. These include data capture and collection methods, materials and instrumentation, data storage sites, data processing and analysis tools, domain standards (formats, ontologies etc.), data sharing practices, databases and publication repositories (e.g. PubMed, Protein Data Bank) and long-term preservation plans. The aim is for bi-lateral learning to occur, with the iSchool students sharing information and guidance e.g. on funder mandates for DMPs, and for the researchers to share their domain-centric data practices. In this example, the immersive unit acts as the Capability Ramp. Evaluation of the courses run to date, highlight a very positive experience for students:

> 'It was great to see a real life example of how a lab generates and uses data.'

'We learned not only about the specifics of their research but also about the lifecycle of data.'

'This was a valuable experience. It was very practical and illuminated some of the struggles that one may encounter in discussing data as its own area of research.'

And for researchers:

'Explaining what one does to a new person is instructive, since it shows you what you do not understand and cannot explain. Discussion with the (LIS) student exposed some weaknesses in my own thinking.'

'Showed them data back up on three drives – they asked a question about fire risk.'

'What happens if fire breaks out or a water pipe bursts – how to log this? If papers are destroyed.'

Partnerships in Practice: Creating a Digital Scholarship Data Observatory

Information professionals working in research libraries have identified and articulated roles in research data management (Lyon, 2007; Jaguszewski and Williams, 2013; Council on Library and Information Resources, 2013), and among academic libraries there is a widespread belief that such roles are appropriate; in a 2012 survey of US academic institutions, 100% of respondents answered "yes" to the question "Do you believe librarians should play a role in managing researchers' digital data?" (Moen and Halbert, 2012). Despite these convictions, libraries face a variety of challenges in developing the organizational capacity to support data management, including a lack of established positions and gaps in understanding the practices and perspectives of researchers and disciplinary data practices (Lyon, 2012).

The University Library System (ULS) at the University of Pittsburgh is representative of both the enthusiasm and the challenges that surround libraries' participation in research data management. In its strategic planning it has identified and committed itself to a broad set of data-related services and capabilities; along with many other research libraries it groups these services under the umbrella term "digital scholarship" (as defined in Smith Rumsey, 2011). However, the ULS has only recently started to develop its own organizational expertise in data curation, and it has lacked dedicated personnel associated with data services or a structured program for building appropriate data-centric knowledge within the organization.

In response, the ULS and the University's iSchool have co-sponsored two postdoctoral research positions specifically designed to catalyse development of data-related services and capabilities within the library, while simultaneously positioning the library as an additional showcase site for the three primary functions of iSchools already articulated: education, research intelligence and professional practice. In this case study, the joint appointment acts as the Capability Ramp and in both instances, the positions have been filled by iSchool doctoral graduates. The partnership between the ULS and the iSchool is designed to emphasize bi-directional relationships and the postdoctoral researchers are fully embedded within both organisations. Reflecting an enabling

diffusion effect, the library staff, iSchool faculty and graduate students are increasingly interacting in and between the two units, resulting in institutional RDM capacity-building, new RDM advocacy programs for researchers and enhanced RDM infrastructure and supporting services.

Mapping Disciplinary Data Practice: Towards Shared Infrastructures and Services

A variety of research methodologies have been applied by iSchool faculty to gather intelligence about disciplinary data practices e.g. interviews (Cragin et al., 2010), surveys (Tenopir et al., 2011), and observational techniques (Wallis et al., 2013). The Community Capability Model Framework[2] (CCM) and the associated Excel-based Profile tool, which can be downloaded from the website, were devised as a mechanism for self-assessment by researchers (Lyon et al., 2012). The developing CCM structure and terminology were informed by a series of international workshops with practising researchers and funder representatives. The CCM provides a rich picture across human, technical and environmental aspects of data-intensive research by investigating eight dimensions in some depth: openness, research culture, common practices, technical infrastructure, collaboration, economic and business, legal and ethical, and skills and training. Building on prior maturity models from the software development field, levels of capability can be measured on a scale of one to five and results data plotted on bar-charts and radar diagrams, facilitating the visualisation of inter- and intra-disciplinary variation. In this case study, the CCM acts as the Capability Ramp. The CCM has been applied to a growing number of domains, including environmental science (DataONE and EarthCube), agronomy, and selected social sciences including anthropology, political science and LIS (Lyon, Patel and Takeda, 2014; Lyon and Jeng, in prep.). There is a Research Data Alliance[3] CCM Interest Group to promote its application and there is considerable scope for wider use of the methodology e.g. across RDA domain-based Interest Groups.

These types of research intelligence collection provide excellent opportunities for iSchools to share new perspectives on disciplinary data behaviors, attitudes and community norms. In addition, as the body of knowledge of different domain data practices grows, there is potential for benefits from enhanced data interoperability across domain silos through emerging consensus on data standards, format types and ontologies, wider adoption of particular tools, protocols and platforms, and ultimately more informed decision-making and strategic investment in data infrastructure by research funding agencies.

iSchool Futures:
Building Potential for the Next Ten Years

This paper has articulated a new Capability Ramp Model as a framework for positioning iSchools as influential and effective agents of change in the data arena. Looking ahead, there are many indicators that the value and importance of data in research, business and society will grow, both in terms of the sheer volume of data to curate and manage (considering the likely ubiquity of embedded sensors, the sophisticated consumer

---

2 Community Capability Model for Data-Intensive Research:
  https://communitymodel.sharepoint.com/Pages/default.aspx
3 Research Data Alliance: https://rd-alliance.org/

marketing of wearables and other mobile devices and the scaling up of observational monitoring and computational modelling/simulations to inform environmental policy and predictive scenarios), and in terms of societal dependence on data-driven information systems in healthcare, commerce, security and defence, to name but three sectors. How should iSchools respond? In conclusion, three steps are outlined, which will help iSchools to realise their full potential value to help to bridge the data talent gap going forward.

As has been noted earlier, an increasing number of iSchools are developing new data-centric programs, courses and specialisations including data curation, research data management, big data analytics, information and data science. Whilst this step is certainly to be welcomed, a further action is proposed whereby such data programs move from specialisations or special topics, to become embedded within the core curriculum and are viewed as central to the mission, or to put it another way, data is mission-critical for iSchools. The dramatic scaling-up of data production from high through-put devices such as sequencers, the large hadron collider and (very) large telescopes, alongside the increasing dependence on the analysis, interpretation and management of data outputs, suggests that something of a seismic shift is happening in the creation and collection phase of the data lifecycle. There now needs to be a parallel transformative re-engineering of data education, training and skills production to keep pace with market demands for data talent.

In recent years much attention has been given to the effective and efficient transitioning of therapies, drugs, diagnostic tools and other treatments, from the research laboratory to the clinical setting and this has been characterised by the term "translational medicine" or more loosely "from bench to bedside". The direct healthcare benefits to patients and wider society are self-evident, but the basic principle is helpfully articulated because it highlights the requirement for organisations to adopt appropriate behaviours and cultural practices in their day-to-day operational practice. The term may also be applied to the data environment and in the specific context of iSchools, where "translational data science" describes the enhanced transition of skills, software tools and intelligence from the iSchool to the marketplace, which may be interpreted as industry, government, libraries, archives or data centers. Adopting a translational perspective will enable iSchools to supply and deploy data talent and data products more rapidly to the range of consumers, where there is currently an acknowledged workforce need.

There are a number of global initiatives to promote data interoperability, data curation and data science; some have focussed on business and industry partners, some are sector-specific, others aim at crossing disciplinary and geographic boundaries. The Research Data Alliance has been very proactive in harnessing global community effort to tackle some of the technical challenges in developing interoperable data infrastructure, standards and workflows. Selected organisations have signed up to become members and much of the work is conducted via community Working Groups, Interest Groups and BOF (Birds of a Feather) Groups. There is an Interest Group for the Education and Handling of Research Data and iSchool representatives have contributed to the semi-annual RDA plenary meetings. However, there remains considerable potential for a more co-ordinated and comprehensive engagement with the RDA and a higher profile for iSchools as a professional body. The benefits to each party are clear; the RDA will gain a rich input from iSchools around education, training, skills, intelligence-gathering and infrastructure development, whilst the iSchools will benefit from collaborating with established community networks, leading data practitioners and domain specialists.

In conclusion, iSchools are key players in the data space. They are already well-connected with primary data stakeholders (e.g. domain researchers, service organisations and the student workforce), they have the ability to deploy Capability Ramps to nurture the much-needed data talent and they have the potential in terms of critical mass, to be transformative in scaling-up the human infrastructure component of the knowledge economy. There will inevitably be challenges ahead, but the next ten years hold much promise for iSchools as influential agents in the world of data.

References

Atkinson, M., de Roure, D., van Hemert, J., & Michaelides, D. (2010). *Shaping ramps for data-intensive research.* Paper presented at the UK eScience All Hands Meeting, 2010, Cardiff. Retrieved from http://eprints.soton.ac.uk/271235/1/DIRampsAHMabstractMPA%2BDDER.pdf

Bakhshi, H., Mateos-Garcia, J., & Whitby, A. (2014). *Model workers: How leading companies are recruiting and managing their data talent.* Retrieved from Nesta website: http://www.nesta.org.uk/sites/default/files/model_workers_web_2.pdf

Borgman, C.L., Darch, P.T., Sands, A.E., Wallis, J.C., & Traweek, S. (2014). The ups and downs of knowledge infrastructures in science: Implications for data management. In *Proceedings of the 2014 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*. doi:10.1109/JCDL.2014.6970177

Callaghan, S., Donegan, S., Pepler, S., Thorley, M., Cunningham, N., Kirsch, P., Ault, L., Bell, P., Bowie, R., Leadbetter, A., Lowrie, R., Moncoiffe, G., Harrison, K., Smith-Haddon, B., Weatherby, A., & Wright, D. (2012). Making data a first-class scientific output: Data citation and publication by NERC's environmental data centres. *International Journal of Digital Curation, 7*(1), 107–113. doi:10.2218/ijdc.v7i1.218

Corrall, S., Kennan, M., & Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends, 61*(3), 636–674. doi:10.1353/lib.2013.0005

Corrall, S., & Pinfield, S. (2014). Coherence of "open" initiatives in higher education and research: Framing a policy agenda. In *Proceedings of the iConference 2014* (pp. 293–313). doi:10.9776/14085

Council on Library and Information Resources. (2013). *Research data management: principles, practices, and prospects.* (CLIR Publication No. 160). Retrieved from http://www.clir.org/pubs/reports/pub160/pub160.pdf

Cragin, M.H., Palmer, C.L., Carlson J.R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A, 368,* 4023–4038. doi:10.1098/rsta.2010.0165

Department for Business, Innovation and Skills. (2013). *Seizing the data opportunity: A new strategy for data capability.* Retrieved from https://www.gov.uk/government/publications/uk-data-capability-strategy

Gray, J. (2009). Jim Gray on eScience: A transformed scientific method. In T. Hey, S. Tansley & K. Tolle (Eds.), *The Fourth Paradigm: Data Intensive Scientific Discovery.* Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf

Heidhorn, P.B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends, 57*(2), 280–299. Retrieved from http://muse.jhu.edu/journals/lib/summary/v057/57.2.heidorn.html

Hey, T., & Trefethen, A. (2003). The data deluge: An e-science perspective. In F. Berman, G.C. Fox, & T. Hey (Eds.), *Grid computing: Making the global infrastructure a reality.* New York: Wiley. Retrieved from http://eprints.soton.ac.uk/257648/1/The_Data_Deluge.pdf

Jaguszewski, J.M., & Williams, K. (2013). *New roles for new times: Transforming liaison roles in research libraries.* Retrieved from Association of Research Libraries website: http://www.arl.org/nrnt

Laney, D. (2012). *The importance of big data: A definition.* Retrieved from Gartner website: https://www.gartner.com/doc/2057415/importance-big-data-definition

Larsen, R.L. (2010). The iSchools. In M.J. Bates & M.N. Maack (Eds), *Encyclopedia of library and information sciences* (3rd. ed.) (pp. 3018–3023). New York: Taylor & Francis. Retrieved from http://d-scholarship.pitt.edu/5852/1/The_iSchools_final.pdf

Larsen, R.L., Palmer, C., Lyon, L., Hedstrom, M., & de Roure D. (2014). *Preparing the workforce for digital curation.* Panel presented at the 9th International Digital Curation Conference, San Francisco, CA. Retrieved from http://www.dcc.ac.uk/webfm_send/1351

Lyon, L. (2007). *Dealing with data: Roles, rights, responsibilities and relationships.* Retrieved from http://opus.bath.ac.uk/id/eprint/412

Lyon, L. (2009). *Open science at web scale: Optimising participation and predictive potential.* Retrieved from Jisc website: http://www.jisc.ac.uk/media/documents/publications/research/2009/open-science-report-6nov09-final-sentojisc.pdf

Lyon, L. (2012). The informatics transform: Re-engineering libraries for the data decade. *International Journal of Digital Curation*, *7*(1), 126–138. doi:10.2218/ijdc.v7i1.220

Lyon L., Ball, A., Duke, M., & Day, M. (2012). Developing a community capability model framework for data-intensive research. In *Proceedings of the 9th International Conference on the Preservation of Digital Objects* (pp. 9–16). Retrieved from https://ipres.ischool.utoronto.ca/proceedings

Lyon, L., & Jeng, W. (in prep.). *Evaluating data-intensive capability: A pilot case study in the social sciences.*

Lyon, L., Patel, M., & Takeda, K. (2014). *Assessing requirements for research data management support in academic libraries: Introducing a new multi-faceted capability tool.* Paper presented at the Libraries in the Digital Age conference, Zadar, Croatia. Retrieved from http://ozk.unizd.hr/proceedings/index.php/lida/article/view/121/123

Lyon L., & Takeda, K. (2012). *What is a data scientist? (Data scientists in the wild).* Presentation at the Microsoft eScience Workshop, Chicago. Retrieved from http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/presentations.html#microsoft-escience-2012-10

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh C., & Byers, A.H. (2011). *Big Data: The next frontier for innovation, competition and productivity.* Retrieved from McKinsey Global Institute website: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Moen, W.E., & Halbert, M. (2012). *Support for research data management among U.S. academic institutions: Results from a national survey.* Retrieved from http://digital.library.unt.edu/ark:/67531/metadc181695/

National Science Board. (2005). *Long-lived data collections: Enabling research and education in the 21st century.* Retrieved from http://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf

Palmer, C.L., Thompson, C.A., Baker, K.S., & Senseney, M. (2014). *Meeting data workforce needs: Indicators based on recent data curation placements.* Paper presented at iConference 2014, Berlin, Germany. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/47308/133_ready.pdf?sequence=2

Royal Society. (2012). *Science as an open enterprise.* Retrieved from https://royalsociety.org/~/media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

Shadbolt, A., Konstantelos, L., Lyon L., & Guy, M. (2014). Delivering innovative RDM training: The immersiveInformatics pilot programme. *International Journal of Digital Curation, 9*(1), 313–323. doi:10.2218/ijdc.v9i1.318

Smith-Rumsey, A. (2011). *New-model scholarly communication: Road map for change.* Retrieved from University of Virginia, Scholarly Communication Institute website: http://uvasci.org/institutes-2003-2011/SCI-9-Road-Map-for-Change.pdf

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PloS ONE, 6*(6), e21101. doi:10.1371/journal.pone.0021101

Wallis, J.C., Rolando, E., & Borgman, C.L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PloS ONE, 8*(7), e67332. doi:10.1371/journal.pone.0067332