

Proceedings of RAILS - Research Applications,  
Information and Library Studies, 2016: School of  
Information Management, Victoria University of  
Wellington, 6-8 November, 2016.

---

## 'In the eye of the beholder': knowledge and skills requirements for data professionals

**Mary Anne Kennan**

**Introduction.** *The professionals required to in data science, data librarianship and data management are a new breed for whom the knowledge and skills are just emerging. Academic institutions are putting together courses to address this shortage without fully understanding what the knowledge and skills requirements are for data professionals in different types of organisation types. This paper aims to increase that understanding.*

**Method.** *Interviews were conducted with thirty-six currently practicing data professionals and their employers about current knowledge and skills requirements. Participants were purposefully selected to achieve a cross section of data roles and employing institutions and interviews ceased once saturation was achieved.*

**Analysis.** *Two types of coding were employed: initial coding to establish categories, followed by more focused coding, for analytical depth. Every effort is made to accurately present the viewpoint of participants.*

**Results.** *All participants reported the importance of high level communication and personal learning skills and characteristics around curiosity, flexibility, and comfort with change. In universities and scientific research organisations the required knowledge and skills are in areas which might be classified as data management and curation; and in business and government organisations, as data science and management.*

**Conclusion.** *While there are still uncertainties about knowledge and skills requirements and role ambiguities in different data roles in different organisational types, knowledge and skills related to particular roles begin to emerge and are discussed.*

### Introduction

It has been proposed that the professionals required to work with data are a new breed for whom the knowledge and skills

requirements are just emerging (e.g. [Davenport &Patil, 2012](#); [Provost &Fawcett, 2013](#)), and in addition there is a limited availability of skilled workers ([Kim, 2016](#)). Organisations need data professionals to enable better use, management, curation, and preservation of data and to explore data reuse, aggregation and sharing. This need is evident in universities and research organisations as well as in business and government organisations ([Australian Government Information Management Office, 2013](#); [Corrall, Kennan, &Afzal, 2013](#); [Tech Partnership, 2014](#)).

It has also been suggested that academic institutions are designing courses to address this shortage without fully understanding the knowledge and skills requirements in different types of organisations for data librarians, scientists, analysts or managers ([Kim, 2016](#); [Provost &Fawcett, 2013](#)). The research reported in this paper investigated what currently practicing data professionals and their employers see as the key knowledge and skills for new professionals coming into data roles in business and government organisations and in universities and scientific research organisations.

## Background literature

The literature is complicated, in that data roles themselves are not clear. While there have been many attempts to classify data roles, there is still no agreement, but some common categories emerge. Swan and Brown ([2008](#)) suggested four main roles each with discrete roles to play in the data lifecycle:

- data creator (people with domain expertise who produce data and may also have a high level of expertise in handling, manipulating and using data)
- data scientist (people involved in creative enquiry, and analysis, of data)
- data manager (computer scientists, information technologists or information scientists who take responsibility for computing facilities, storage, continuing access and preservation of data)
- data librarian (people trained and specialising in the curation, preservation and archiving of data).

Lyon and colleagues ([Lyon and Brenner, 2015](#); [Lyon, Mattern, Acker and Langmead, 2015](#)) proposed slightly different families of data science roles and also suggested their typical organisational locations:

- data analyst (people conducting business/scientific analytics, mathematics, statistics, modelling; corporate sector)
- data archivist (people involved in long term preservation, repository management; national archive)
- data engineer (people involved in software development, coding, programming, tools; information technology company)
- data journalist (people using data to tell stories and providing news using visualisations; newspaper publisher)
- data librarian (people involved in advocacy, research data management, training; university or research institute)
- data steward and/or data curator (people involved in curation, cleansing, annotation, selection and appraisal; data centres)

By [2012](#) Stanton, Palmer, Blake and Allard discussed the concept of a *T-shaped professional* for data professionals, where broad data knowledge would be complemented by depth of knowledge in one of three areas: data curation; analytics, visualisation, preservation; networks and infrastructure. An *I-shaped* model was also proposed, which included domain knowledge at the base (Harris, Murphy and Vaisman, [2013](#)).

For universities and research organisations, Cox and Pinfield ([2014](#)) suggested that when the Digital Curation Centre's lifecycle model is addressed, data management seems more related to archival and records management than librarianship, although they did map existing library roles to required data competencies. There is an increasingly vast literature that addresses evolving data roles and the education and skills required to work as data managers and data librarians in universities and other scientific and research organisations (*c.f.* Bailey, [2017](#)). The literature is complicated by related considerations of the education and training needs of researchers, in addition to that of those such as data librarians or data managers potentially involved in curating and/or managing data created by others ([Carlson, Johnston, Westra, and Nichols, 2013](#); [Friedlander and Adler, 2006](#); [Henty, 2008](#); [High Level Expert Group on Scientific Data, 2010](#); [Lyon and Brenner, 2015](#); [Molloy and Snow, 2012](#); [Pryor and Donnelly, 2009](#); [Swan and Brown, 2008](#)). [Lewis \(2010\)](#) suggested that the focus of data roles for library staff lies in developing the data literacy and data management skills of both graduate research students and undergraduates. Others suggested the foci of data librarians should include (among other things) advocacy and the training of data creators in research data management (Lyon and Brenner, [2015](#)),

or adding metadata to data, and preparing data for curation (Li, Xiaozhe, Wenming, and Weining, [2013](#)).

Participants in a 2012 survey recognised knowledge and skills gaps in university libraries in data curation, information and communication technology, subject and/or disciplinary knowledge, research methods and processes ([Corrall et al. , 2013](#)). While there may currently be skills gaps, Witt ([2012](#)) suggested that much of this work will just become part of regular library practice, whereas others noted there may be a continuing need for a series of specialists (Kim, Addom, and Stanton, [2011](#)), although there may be some overlapping roles (and correspondingly, knowledge and skills) ([Lyon et al., 2015](#)).

In business and government, there is a similar lack of clarity about the particular knowledge and skills required for data science roles and as [Harris, Murphy and Vaisman \(2013, p.3\)](#) suggest, this can lead to impaired communication and difficulty in matching '*talent to projects*'. For data scientist and data analyst work, generally a wide range of knowledge and technical skills are required, from computer science, to analytics, mathematics and statistics, to modelling and to domain expertise ([Bertolucci, 2013](#)). While organisations might like to recruit people with all these skills, they rarely reside in one person; hence the description of such people as the mythical data science unicorn ([Bertolucci, 2013](#); [Ramanathan, 2016](#)).

For universities deciding what to teach in their data-related courses, and students deciding what subjects would be useful in their data careers, there are still many questions. In terms of translating roles from their job advertisement analyses into curricular recommendations, Kim *et al.* ([2011](#)) propose a list of the top ten required courses which include: data curation, database design and management, project management, essentials of scientific research, overview of cyberinfrastructure, geographically distributed collaboration, Web management and design, scripting or introductory programming, data mining and information systems.

While it is relatively common in the United States ([Harris-Pierce and Quan Liu, 2012](#); [Kim, 2016](#)) for iSchools and library and information science (LIS) schools to offer data courses, including courses on data curation and data management, it is less common in Australia. In Australia, however, there are courses in data science and data analytics, although they generally focus on the

business and technological aspects of data, rather than the information or data management and curation aspects of data. (See the list compiled by Raylee Macaulay and the author and shared through the *ANDS 23 (research data) things* Website under the heading *Thing 23: Making connections* and the subheading [Get a data qualification](#).

While there is recognition of a skills gap, there is also discussion in the literature about whether formal training as part of library and information studies, information technology or business degrees, or even a new degree in data science is appropriate. Others believe that learning on the job or through continuing professional development is most appropriate ([Corrall et al., 2013](#); Cox and Pinfield, 2014; [Cox, Verbaan, and Sen, 2014](#); [Harris, Murphy and Vaisman, 2013](#)). The author worked in an institution that was considering offering a course in data management and thus sought to understand directly from professionals working in data roles what they felt are the knowledge and skill requirements for current and future data professionals (see Appendix 1: Semi-structured interview schedule). To develop that understanding the following questions were asked:

1. Where, and in what roles, are data librarians, managers and scientists likely to be employed?
2. What knowledge and skills are required of people working as data professionals?
3. What are the educational and training requirements for data professionals?

## Method

In 2015 thirty-six practicing data professionals and their employers or supervisors in Australia were interviewed from universities, scientific, government and business organisations (see Appendix 3 for details). The interviewees were located in libraries, research offices, information and/or data management departments, and business strategic or analytic units, and information technology (IT) units. Participants were purposefully selected in order to achieve a cross section of data roles and interviews ceased once saturation (the point in data collection when no new or relevant information appears to emerge with respect to the topic under investigation ([Saumure and Given, 2008](#))) was achieved. It was evident after the first few interviews that data work is often conducted by teams of differently qualified professionals, particularly in the larger organisations, rather than by individuals

or groups of similarly qualified people and this is reflected in the literature ([Sands et al., 2014](#)). In such cases where a group of differently qualified people were working together for a common purpose (supporting data management in their organisation) group interviews were held, so that all involved could contribute and be aware of what was being said. The interviewer took care to ensure that all present were given the opportunity to respond to each question or talking point ([Cohen, Manion and Morrison, 2013](#)). In other cases, one person, such as the originally contacted data professional or a person with a management role, spoke about their own role and the roles of their team members.

As the research progressed it also became evident that the knowledge and skills reported as required in business and government organisations were quite different and more aligned with what has become known as data science than data librarianship or management. Accordingly, participant quotes are attributed to the organisation type: scientific research organisations; universities (mainly but not exclusively libraries); financial institutions; government departments; and the utility company. Selection was based on participants having, or supervising people with, a data role and thence on participant availability. Interviews continued until content saturation was achieved. Individual interviews were often conducted via Skype. All group interviews were conducted face-to-face. The study was approved by the Ethics Committee of Charles Sturt University. Participation in the study was voluntary and all participants had the study explained to them.

Interviews were semi-structured, audio-recorded and transcribed using a professional service. Two types of coding were employed: initial coding to establish categories such as roles, knowledge and skill requirements, education and training requirements; followed by more focused coding looking at recurrent patterns, interconnections and variations, for analytical depth. The focussed coding informs the findings presented below, with quotes from respondents to assist the reader with understanding the trustworthiness of the analysis and reporting. Every effort is made to accurately present the viewpoint of participants, through the use of quotations and examples.

## Findings

### Where, and in what roles, are data professionals likely to be employed?

In universities and research organisations participants reported that they and their fellow data professionals worked in the libraries and other *information* departments such as information technology (IT) and information management and research offices of universities and scientific organisations. Job titles and work roles differ across organisations. There was a huge range of job titles for those working in data roles. While many reported roles had the word data in the title (data librarian, data manager, data specialist), occasionally the role was titled something more general such as eresearch librarian, eresearch manager, or even project officer (for examples see Appendix 1, column 1). In most of the universities and larger scientific organisations participants reported that it was not just data librarians and other data specialists who need to understand data, but most library staff, whether they are subject liaison, outreach, research or other librarians: *'It's part of our library plan I believe, so strictly speaking we are all supposed to take an interest ...'* (university).

Others, both in universities and in scientific research organisations, talked about the importance of data literacy not just for librarians working with researchers, but with the researchers themselves:

*I really wonder if it should be more about trying to embed better information and data management skills in the curricula of other professions ... shouldn't there be some kind of information management and data management compulsory training for anyone that's going on to do any kind of advanced research in a university environment.*  
(university)

Several respondents mentioned that there is no one size fits all:

*... the data librarian in another environment might be a completely different role altogether. When I go out and talk to people doing a data librarian role they're not doing any of what I'm doing.* (university)

In business and government organisations, data professionals also work in a range of roles across the organisation. One role was a data analytics role, explained as roughly equated with what is called in the press and literature data science or data analyst roles and usually in the business side of the organisation. The other major tranche of roles was in information technology (IT) departments in roles such as information and data architect, information and data manager, data modeller, or database

administrator. However, as a respondent from a financial institution noted, these are just *'typical roles [and] they vary from organisation to organisation, even within the organisation'*. As well, the data scientists and data analysts were reported as being in higher demand if they had programming skills in addition to analytics skills. This was similar in the government organisations, but there was a third kind of role, information managers, who came from either an information management, information science, information technology, or archives and records background, who are often tasked with a bridging role, looking after the data and bringing the information technology and analyst and data scientist roles together.

So generally speaking data roles are

*not clearly defined, and the reason why –well ...one reason ...is because the data is so fragmented that we get it from so many bits and you go to so many areas that it's hard to say who owns data across all the fragmented bits and when you look at the organisation is structured into its silos and sub-silos data doesn't fit nicely into those silos because it is pervasive and it does stretch across a lot of those silos. (financial institution)*

The next section will look at what interviewees say data professionals are actually doing and the knowledge and skills required to do the work.

## What knowledge and skills are required of people working in data roles?

Interestingly, in almost all the organisations respondents indicated that the major set of skills they required, or that they were looking for in data librarians and others in the data field, were more generic than field specific and these could be labelled interpersonal skills and behavioural characteristics, *'... looking more at attributes than at concrete knowledge or skills because I want people who are flexible ...'* (university).

## Interpersonal skills and behavioural characteristics

In all types of organisations, the most commonly mentioned interpersonal skills were high level communication skills, such as advocacy, negotiation, and capability building skills and including writing of documentation, use-cases and other technical writing. In terms of personal characteristics, participants reflected that people

working in the data space need to be comfortable with change, have a service philosophy, willingness to learn, discretion, *'boundless curiosity'* and be adaptable, assertive, and open to new experiences. The ability to network and use networks to learn and keep up to date were also frequently mentioned. Several participants also mentioned the importance of being confident enough to *'know what you don't know'* (scientific research organisation) and when it is time to call someone with different knowledge or skills in. In the words of one participant: *'Human skills, being able to liaise with fellow professionals'* (government department). Participants from the financial institutions were more specific and stressed the relationship of the human skills in relation to the context of the organisation and the data or information, for example:

... [the importance of the] notion of systems being people and systems and processes. So, understanding the use of data and information within its context means you've got to understand and work with the audience; means you've got to understand and work with the receivers of that information; means you've got to understand the consumers of that information. (financial institution)

## Contextual and domain knowledge

In the research and scientific organisations almost all participants mentioned the importance of contextual knowledge about the research environment of the university or scientific organisation, and related funding agency policies, research measures and research evaluation activities. In addition to organisational context, the importance of understanding, or being open to, discipline specific research life cycles and cultures, processes, ethics, disciplinary research methods and scholarly communication mores (e.g. attribution, publishing preferences and citation), intellectual property and licencing laws and policies, access norms, and cultural sensitivities. Particularly the librarians and other staff in the scientific research organisations felt that domain scientific knowledge was either a prerequisite or something useful that should be learned on the job and several of the universities also mentioned that subject speciality in specific areas would be encouraged.

In the business and government organisations, the importance of context was also emphasised, but of course, the context is different, for example ideally people managing and making decisions about data should know:

*... what does[the] organisation do, who are its customers and what is their context. And what are the regulations [that] apply to the business as well ... industry regulations, we have the [different] country regulations ... so say for example if we run, if an organisation runs exactly the same business overseas in Singapore, there would be different regulations for the organisation's data in the two parts of the same organisation. (financial institution)*

>

However, the business and government organisations were clearer about how they would not expect new graduates to know all contextual knowledge, but just to be aware that the context in which data and information are used are highly varied, understand examples, and where to look for specific contexts and be prepared to continue learning on the job, for example:

*If I take on a university graduate, I need them to understand the concepts, the underlying general knowledge about data and information and then [they can learn] the specific and practical examples on the job. (financial institution)*

*You don't have to know, say if you're going to Department of Health you don't expect to know everything what's there on health, all the taxonomies but as a data management professional and information management professional you should be able to know that there are standards that each industry has ... ; for each industry there's best practices; there's legislation and ... you don't have to know everything yourself but you have to know where to find it and how it all connects... (government department)*

*... you can come in with all the qualifications in the world but the other thing is that in most organisations the information or data management or the information governance rules can be extraordinarily varied in terms of what you're actually called upon to engage in ... and I think you really need someone who is engaged and adaptable and can perform across those ... (government department)*

## Data specific knowledge and skills

Knowledge, or willingness to learn, specific knowledge in the data domain and related data skills were also considered very important. However, while there were some similarities there are also major differences by organisation. Scientific research organisations emphasised the need to understand the variety of

data from flat textual, to relational, numerical, instrument generated data, archival and cultural heritage data and the need for data to be active for long periods and/or constantly updated (for example meteorological, historical, oceanographic, geoscientific and other *'really big data sets'*). In universities and scientific organisations several sites mentioned learning analytics and student data as research data and a couple mentioned physical data such as cassettes, other tapes, papers, photographs, bark paintings, *'frozen fish bits'*, images, and text as data that needed to be dealt with by their data librarians and other data staff. In business and government departments data analytics roles were already very common, but there is a need for more, and the roles are changing from quantitative analytics, to text analytics and data mining. A greater need was also expressed for visualisation skills, to be able to take data and make it meaningful for managers and decision makers in the organisation, and beyond the organisation for customers and other stakeholders.

Frequently mentioned in all organisations was data that may be used in multiple contexts, and over time and therefore the need for data flow from one organisation or organisational unit to another and currently there are a dearth of people who understand this and

*... what the data or information is and will be actually used for. So, it will be gathered in one area but re-used in seven or eight and quite often there is a disconnect between the people that gather that information and throw it into the ... system ... because they don't have to re-use it so the disconnect between the gathering and the usage creates an issue. And there are also people who will use that particular set of data without any understanding of how it's used elsewhere. (government department)*

In addition to understanding and learning how to manage and use different data types, also mentioned were the need for knowledge, and the related skills to facilitate data sharing, linked data, the data management lifecycle, data management processes such as quality control, data processing, data management planning, and an ability to understand and support data storage requests. Data professionals should also be aware of the legal and regulatory frameworks relevant to data in different contexts. Issues specifically mentioned in universities and research organisations were ethics and consent, copyright and creative commons issues, while the customer and regulatory environment was most mentioned in business and government organisations.

In the universities and scientific research organisations, there was

some mention of the need for data analysis knowledge, mainly as an emerging need in universities, although in the scientific research organisations it was mentioned as a specialised need (as in the business organisations), particularly as a specialised role assisting scientists when data from two or more sources was to be merged and analysed. Data mining as a useful skill was only mentioned by two of the participants in universities and scientific research organisations. By contrast these skills and related knowledge were strongly expressed need in business and government organisations, more frequently required now than even five years ago. Data analysts in the past needed a background in statistics, science, engineering, or mathematics. Now they need that, but also text analytics and one of the financial institutions talked about their new team where in addition to these skills they had a new team called the analytics and information team: *'So these guys, a lot of them are [graduates in] business studies, economists, they've got economics degrees'* and both the data analyst and the IT professional interviewed at this organisation held a master's in business and one was undertaking a PhD as well. What was interesting was when participants described what these people actually did in terms of work; their job was to make the information contained in data accessible to non-data professionals, such as managers:

*There's no point in going to them [management] with a spreadsheet or a table with 40 numbers in it, because it means nothing. So, you've got to actually be able to translate, it's an ability to translate and interpret from the detail into something that's you know, you can consume, like a lay person. (utility company)*

*... takes data and makes pictures out of it that management can understand. (financial institution)*

Interestingly only two university participants explicitly mentioned *'data curation'* as important, although others did mention aspects of curation, such as the importance of understanding of the data lifecycle and data archiving, provenance, the importance of data as ongoing records, digitisation of analogue data, and digital preservation. When asked about curation at the end of the interviews, university participants said that they didn't mention it because at the present time curation is *'rare because of costs'* and also because of costs respondents felt they *'can't make many assurances'* about preservation although they *'would like to do more'*. It seems that at the moment the focus is more on

description and storage and the time for curation, including preservation, is yet to emerge: '*... curatorial leadership. It's absent presently and it's a real gap*' (scientific research organisation).

Almost every participant mentioned the importance of knowledge about, and skill in applying, metadata. Specific terms used included '*advanced metadata*', '*metadata standards*', '*metadata mapping*', '*cataloguing*', '*structured information*', '*ontologies*', and '*metadata harvesting*'. Different terms are used in different contexts to define the structured description of data, enabling access, use and re-use. While specific and different metadata standards and schemas were mentioned in almost every interview and frequently, (for example Dublin Core, Darwin Core, RDF, DIF, ANZLIC, RIF-CS) most participants emphasised that knowledge about the purpose and use of metadata and a willingness to accept and learn the different standards and schemas applicable in different contexts was necessary.

While one of the purposes of metadata is to provide access to the data or information described, knowledge and skills about enabling access in other ways were also frequently mentioned by participants. For example, the importance of knowledge about '*discovery (how people look for information)*' was mentioned in both universities and scientific research organisations, as was more technical access knowledge and skills, such as knowledge about the user experience, Web interfaces, and graphical user interfaces (GUIs). The dynamic nature of data was frequently mentioned in the access context, so understanding of digital object identifiers (DOIs), back-up, version control and naming conventions were also listed as important.

One interesting phenomenon emerging, particularly in the business and government organisations, was the way in which words, data and information were used interchangeably. There was an awareness that data was more than the *bits and bytes* of the definitions of earlier times in business, which warrants further investigation.

## Information technology

Many issues related to access to, and storage of, data are managed in the information technology (IT) realm, for example, Web interfaces and graphical user interfaces. While important IT knowledge and skills were raised (discussed below), most participants said that while IT skills were important for data

professionals, it was more important to have *'just enough'* of an understanding of IT, to be able to 1) bridge the perceived communication gap between IT departments and researchers, managers and other professionals, and 2) understand the IT options and make informed decisions, rather than to be able to actually *'do'* IT. For example, one participant (scientific data coordinator) had overseen the development of the data centre Website in his organisation, but not actually done the Website development. In the words of another participant, IT knowledge and skills are *'not strictly necessary, but it helps'* (scientific research organisation). Specific useful IT knowledge and skills that were mentioned as useful to have were XML, database structure and design, APIs, user centred design, natural language processing tools, sensor networks, and the internet of things. Programming was also frequently mentioned, but different languages in different contexts and, again, the importance of being *'programming savvy'* (scientific research organisation) was stressed rather than the requirement to be an actual programmer or be an expert in a particular programming language. Examples of useful programming skills were mentioned though and included Python (scientific research organisations, universities, government departments, financial institutions) SPSS (universities), SQL (financial institutions, scientific research organisations), Java (universities, financial institutions), Unix (financial institutions), and C++ (financial institutions). Useful systems and tools for data professionals to understand and use included Web services (scientific research organisations), open journal systems (universities), open formats (universities), collaborative tools, visualisation tools, and analysis tools (universities, scientific research organisations). Less frequently, but still occasionally, mentioned were infrastructure, software development, machine coding, software coding, technical and use-case writing and database development and management skills.

The higher order IT management related knowledge and skills of information architecture and information governance were only raised by the financial institutions, government departments and two of the scientific research organisations, and interestingly not by the universities. However other management knowledge and skills requirements were frequently mentioned by all participants, especially project management, but also stakeholder and relationship management, and change management. One academic librarian specifically mentioned the need to understand and create policy. Mentioned by fewer participants, but still important in all

types of organisations, were strategic planning, internal consulting, business analysis, and stakeholder requirements gathering.

## Training and advocacy

Related to communication skills, is the important work of data managers and librarians, particularly in the universities, but also in the larger research organisations, to put together workshops and presentations around legal and regulatory frameworks, copyright and creative commons, service availability, policy (both organisational and external), processes (such as version control, back up, naming conventions), and storage, and to provide data literacy workshops to higher degree by research students and early career researchers. Advocacy and training were not mentioned by the financial institutions, utility company or government departments in their role for data professionals, except regarding training that their organisation would offer data professionals. This appears to be a big difference in data roles, highly important in universities and research organisations but not mentioned in business and government organisations.

## Teamwork

While participants were talking about data knowledge and skills that might be useful to data professionals in their organisations, most also recognised that not all these knowledge and skills might currently reside in the one person or role, particularly in the larger organisations. Instead, particularly in larger organisations, teams of people provide data support and they may reside in the library, in IT, or another part of the organisation. As one respondent said, *'You tend to bring together teams of different specialists'* (scientific research organisations).

The teams might consist of metadata, data and/or information management, IT, analysis, storage, and other specialists who work as a data support team to bring their organisations a wide variety of data services. Teams may not be specifically designated data teams, but there may be data professionals in many different parts of the organisation who need to work together to generate, manage and use the organisation's data. Thus, teamwork was recognised as critical. As stated: *'One person cannot do all this'* (government department).

## What are the educational and training requirements for data professionals?

With regard to the education and training of those currently working in a data manager or data librarian role, or for whom data was a significant part of their role, in universities, scientific research organisations and government departments, most had a background as librarians or information managers, some also with prior qualifications in IT, and several had IT only backgrounds. These people had learned largely '*on the job*' data specific knowledge and skills on top of their previous professional qualifications. However, in three of the scientific research organisations the data professionals had scientific backgrounds (PhDs, although one said that for data work a bachelors should be sufficient) and had learned their data and information management knowledge and skills also '*on the job*'. The most common ways of learning on the job were directly from colleagues, '*learning by doing*', and through formal professional development such as webinars, conferences and workshops.

*Well I think I'm a common example in the research data management space of a hybrid. We have hybrid professionals ... who ... bring absolutely everything together that you've acquired sometimes through education and sometimes through experience, sometimes by taking a leap of faith or watching other people do it. (scientific research organisation)*

When asked to look towards the future and speculate about what might be the ideal educational and training requirements for data managers and librarians, there was a wide range of responses. Most commonly mentioned were either '*librarians with more*' or '*scientists with more*', where in some cases the '*more*' included data specific skills and in other cases IT education or training. Many of the respondents in the scientific research organisations felt it was still important for data professionals to have the relevant scientific qualification (usually a PhD) backed up with formal data education or training. It was also recognised that in some cases there was a need for formal external courses, for example, in project management, data analysis, XML and other IT, whereas there was also a place for internal training on internal systems, policies and procedures. While some participants felt that short or mini courses were '*the way to go*', others felt that either a specialisation in a library and information studies masters or a graduate certificate would be appropriate.

*Yeah, and so, I guess, you know, if I had to summarise all that in one sentence it's, basically that a narrow focus on librarianship would prepare you less for this than a more*

*generic kind of qualification that covered aspects of archives and knowledge management type stuff, as well, because it's definitely bits from all of those that are really useful. (university)*

With regard to the education and training of those currently working in data roles in the business organisations and increasingly in government departments, there was more emphasis on the knowledge and skills and thus the education for those who the literature calls data scientists. There was an emphasis on business and management qualifications, which would need to include statistics, mathematics and other analytics, including text analytics. However, equally important were IT and programming skills, with computer science and computer engineering mentioned as the background of many in data teams, with no lessening in these requirements. At this stage, it seems that generally these two sides (analytics and IT) of a business work separately and come together on data projects, but increasingly are working together in data teams. In the words of one of the IT people:

*So, these guys, a lot of them are business studies, economists, they've got economics degrees. I have looked at them, some of them are outstanding programmers as well. So, a couple of them have made a system, if you look at it you'll..., I'm impressed and I'm in IT. (financial institution)*

Several people mentioned that some understanding of data and the role of data in business, research and other organisations, should be a part of every degree:

*I think it should be part of every degree just like I think communications should be part of every degree ... data and information so it's at the heart of every education I think. So essentially every graduate in my mind should have a good broad understanding of data and information. Those that are going to be walking into roles where they have to define data and be clearer about it should actually go deeper in that than others. (financial institution)*

## Discussion and conclusion

The interviewees worked mainly in libraries, data centres, and information technology or management sections of their organisations. One of the participants from the financial institution and the participant from the utility company worked on the business side as analysts and one of the financial institution technologists aimed to be a data analyst or data scientist. In the

universities, the reported roles had a range of titles and included knowledge and skills which Swan and Brown (2008) suggested fall within both the Data manager and Data librarian roles, and within all three more specialist roles (Data Librarian, Data archivist and Data steward) suggested by Lyon *et al.* (2015), even when they had the title Data librarian. This is perhaps because of the nature of the work in the organisations in which participants worked, although these previous authors also acknowledge overlap between roles.

Knowledge and skills required are therefore around data librarianship, management and curation. Scientific research organisations had similar requirements to the universities, but were more likely to require also scientific domain knowledge and qualifications. In research organisations, often (but not always) data managers had a background as data creators (scientists and other researchers). Government departments had similar requirements to universities, but instead of scientific domain knowledge increasingly saw need for analytics and modelling. The business organisations required people with data analytics skills, but increasingly saw a need for people with both analysis and computer science and technology skills.

The emphasis of practicing professionals from all organisational types on interpersonal skills and behavioural characteristics is interesting from an educator's point of view. Some attributes reported within the communications skills category, such as technical writing, can be taught and the importance of advocacy, negotiation and capability building can be emphasised and practiced. Indeed, these attributes are part of many existing library and information studies programs (Lyon *et al.*, 2015; Pryor and Donnelly, 2009). However, others such as willingness to learn, discretion, curiosity, and adaptability are more likely to be individual personality traits, which although they maybe encouraged and developed by education, are not always 'teachable'. Further enquiry is warranted to investigate how these attributes may be built into appropriate educational programs. The required knowledge and skills reported by the data professionals in universities and scientific research organisations in this study generally reflect those reported in the literature, with a lesser emphasis on data curation, which is however, noted as a future requirement. This may reflect a difference in the organisational context in Australia, as Harris-Pierce and Liu (2012) noted data curation as a mainstay of data courses in the United States. Some data specific knowledge and skills reported here can be

incorporated into existing library and information studies curricula, for example, and others will need new subjects or courses to be written.

In short, there are identifiable roles, with associated identifiable knowledge and skill requirements related to those identified in previous studies, but also slightly different, which may possibly relate to national differences, or changes in roles as data work matures and there is more understanding by organisations and individuals of the roles and associated knowledge and skills requirements for data professionals. The importance of team work and knowledge and skill sharing was constantly emphasised by all the participants from all organisations. Table 1 represents an effort to classify data roles at this point in time, as uncovered in this project.

	Data librarian and/or Data manager (universities, scientific research organisations, government departments)	Data information technology and systems experts (all participating organisations)	Data scientist (business organisations, government departments)	Data creator (universities, scientific research organisations)
<b>Required</b>	Data management, including description, planning, storage, preservation, archiving. Legal and regulatory frameworks. Data description, metadata. Training and advocacy skills. Knowledge of a wide variety of data formats.	High level programming, software development. Data and information governance. Data description ontologies and modelling. Development and use of software tools and applications.	High level mathematical and statistical analysis skills. Text and other qualitative analytical skills. Data mining, modelling and visualisation. High level of understanding of the business or organisation and its environment.	Domain knowledge, generation of data, expertise in handling, manipulating and using data.
<b>Desirable</b>	Domain and contextual knowledge. Understanding of programming and app	Contextual knowledge. Knowledge of business environment.	Programming, software and app	Data management, data

	development. Data curation. Data visualisation and some data analysis skills	Know ledge of regulatory environment	development.	management planning.
--	---	--	--------------	-------------------------

**Table 1: Data roles with knowledge and skill requirements**

Despite this developing classification, all interviewees recognised that the knowledge and skills required to work as a data librarian or data manager are varied and come from a variety of disciplinary backgrounds; in some organisations, *'librarians with more'* are filling the role and in other organisations *'scientists with more'*. There is no data management or data librarian unicorn, just as there is no data science unicorn, at the moment different organisations have different requirements, or if they do exist they are very rare (Harris *et al.*, 2013; Ramanathan, 2016). Different roles perform different aspects of data work that all need to be brought together in the organisation context, hence the requirement for team work and advanced communication skills in all data roles. As data work becomes more common and more visible in organisations, it is likely that the differences between the different roles will be clarified, providing less role ambiguity and making it easier for employers to identify clearly their data needs; professionals and students to strengthen and explain their areas of interest and expertise; and universities to design programs suited to particular data roles.

## Acknowledgement

The author would like to thank Raylee Macauley for assistance in compiling the first draft of the list of data courses in Australia made available through the ANDS 23 (research data) things Website under the heading *Thing 23: Making connections* and the subheading [Get a data qualification](#).

## About the author

Dr Mary Anne Kennan is Associate Head and Senior Lecturer in the School of Information Studies at Charles Sturt University, where she is also the Higher Degree by Research Coordinator. Mary Anne's areas of research and teaching include scholarly communication and open access, research data sharing and management, library and information studies education in Australia, and the role of information (access and services) in social inclusion. She can be contacted at [mkennan@csu.edu.au](mailto:mkennan@csu.edu.au)

## References

- Australian Government Information Management Office. (2013). [The Australian public service big data strategy](http://www.finance.gov.au/sites/default/files/the-australian-public-service-big-data-strategy-archived.pdf). Retrieved from <http://www.finance.gov.au/sites/default/files/the-australian-public-service-big-data-strategy-archived.pdf>. (Archived by WebCite ® at <http://www.webcitation.org/6svzGcSqq>)
- Bailey, C. W. (2017). [Research data curation bibliography](http://digital-scholarship.org/rdcb/rdcb.htm). Retrieved from <http://digital-scholarship.org/rdcb/rdcb.htm> (Archived by WebCite ® at <http://www.webcitation.org/6svzWkcdI>)
- Bertolucci, J. (2013). [Are you recruiting a data scientist or a unicorn?](http://www.informationweek.com/big-data/big-data-analytics/are-you-recruiting-a-data-scientist-or-a-unicorn/d/d-id/899843) *InformationWeek*. Retrieved from <http://www.informationweek.com/big-data/big-data-analytics/are-you-recruiting-a-data-scientist-or-a-unicorn/d/d-id/899843>. (Archived by WebCite ® at <http://www.webcitation.org/6svzoFgZY>)
- Carlson, J., Johnston, L., Westra, B., Nichols, M. (2013) Developing an approach for data management education: a report from the data information literacy project. *International Journal of Digital Curation*, 8(1), 204-217.
- Cohen, L., Manion, L. and Morrison, K. (2013). *Research methods in education*. London: Routledge.
- Corrall, S., Kennan, M. A. & Afzal, W. (2013). Bibliometrics and research data management services: emerging trends in library support for research. *Library Trends*, 61(3), 636-674.
- Cox, A. M. & Pinfield, S. (2014). Research data management and libraries: current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299-316.
- Cox, A. M., Verbaan, E. & Sen, B. (2014). A spider, an octopus, or an animal just coming into existence? Designing a curriculum for librarians to support research data management. *Journal of eScience Librarianship*, 3(1), 15-30.
- Davenport, T. H. & Patil, D. (2012). Data scientist. *Harvard Business Review*, 90, 70-76.
- Friedlander, A. & Adler, P. (2006). [To stand the test of time: long-term stewardship of digital data sets in science and engineering: a report to the National Science Foundation from the ARL workshop on new collaborative relationships: the role of academic libraries in the digital data universe, Arlington, Virginia, September 26-27, 2006](http://files.eric.ed.gov/fulltext/ED528649.pdf). Retrieved from <http://files.eric.ed.gov/fulltext/ED528649.pdf> (Archived by WebCite ® at <http://www.webcitation.org/6sw0RP8S2>)

- Harris, H.D., Murphy, P.H. & Vaisman, M. (2013). [Analyzing the analyzers](http://www.oreilly.com/data/free/files/analyzing-the-analyzers.pdf). Sebastopol, CA.:O'Reilly Media. Retrieved from <http://www.oreilly.com/data/free/files/analyzing-the-analyzers.pdf> (Archived by WebCite ® at <http://www.webcitation.org/6sw0bjNAT>)
- Harris-Pierce, R. L. & Quan Liu, Y. (2012). Is data curation education at library and information science schools in North America adequate? *New Library World*, 113(11/12), 598-613.
- Henty, M. (2008). [Developing the capability and skills to support e-research](http://www.ariadne.ac.uk/issue55/henty). *Ariadne*, 55. Retrieved from <http://www.ariadne.ac.uk/issue55/henty> (Archived by WebCite ® at <http://www.webcitation.org/6sw1AyRgZ>)
- High level Expert Group on Scientific Data. (2010). [Riding the wave: how Europe can gain from the rising tide of scientific data](http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=62044). Retrieved from [http://ec.europa.eu/information\\_society/newsroom/cf/itemlongdetail.cfm?item\\_id=62044](http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=62044) (Archived by WebCite ® at <http://www.webcitation.org/6sw2QUte6>)
- Kim, Y., Addom, B. K. & Stanton J. M. (2011). Education for eScience professionals: integrating data curation and cyberinfrastructure. *International Journal of Digital Curation*, 6(1), 125-138.
- Kim, J. (2016). Who is teaching data: meeting the demand for data professionals? *Journal of Education for Library and Information Science*, 57(2), 161-173.
- Lewis, M. (2010). Libraries and the management of research data. In S. McKnight (Ed.), *Envisioning future academic library services: initiatives, ideas and challenges* (pp. 145-168). London: Facet.
- Li, S., Xiaozhe, Z., Wenming, X. & Weining, G. (2013). The cultivation of scientific data specialists: development of LIS education oriented to e-science service requirements. *Library Hi-Tech*, 31(4), 700-724.
- Lyon, L. & Brenner, A. (2015). Bridging the data talent gap: positioning the iSchool as an agent for change. *International Journal of Digital Curation*, 10 (1), 111-122.
- Lyon, L., Mattern, E., Acker, A & Langmead, A. (2015). [Applying translational principles to data science curriculum development](http://d-scholarship.pitt.edu/27159/). In *iPres 2015, November 2-6, 2015, Chapel Hill, North Carolina*. Retrieved from <http://d-scholarship.pitt.edu/27159/> (Archived by WebCite ® at <http://www.webcitation.org/6sw1WqkmP>)
- Lyon, L. & Mattern, E. (2016). Education for real-world data science roles (part 2): a translational approach to curriculum development. *International Journal of Digital Curation*, 11(2), 13-26.
- Molloy, L. & Snow, K. (2012). The data management skills

- support initiative: synthesising postgraduate training in research data management. *International Journal of Digital Curation*, 7(2), 101-109.
- Provost, F. & Fawcett, T. (2013). data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51-59.
- Pryor, G. & Donnelly, M. (2009). Skilling up to data: whose role, whose responsibility, whose career? *International Journal of Digital Curation*, 4(2), 158-170.
- Ramanathan, A. (2016, November 21). [The data science delusion](#) [blog post]. Retrieved from <http://datasciencecentral.com/m/blogpost?id=6448529%3ABlogPost%3A489994> (Archived by WebCite® at <http://www.webcitation.org/6sw1j9mkw>)
- Saumure, K. & Given, L. (2008). Data saturation. In L.M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 195-196). Thousand Oaks, CA: Sage.
- Sands, A. E., Borgman, C. L., Traweek, S. & Wynholds, L. A. (2014). We're working on it: transferring the Sloan Digital Sky Survey from laboratory to library. *international Journal of Digital Curation*, 9(2), 98-100.
- Stanton, J., Palmer, C.L., Blake, C. & Allard, S. (2012). [Interdisciplinary data science education. Special issues in data management](#). In *American Chemical Society Symposium Series 1110*, 97-113. Retrieved from <http://pubs.acs.org/doi/10.1021/bk-2012-1110.ch006> (Archived by WebCite® at <http://www.webcitation.org/6sw2pmIvu>)
- Swan, A. & Brown, S. (2008). [The skills, role and career structure of data scientists and curators: assessment of current practice and future needs](#). Retrieved from <http://www.jisc.ac.uk/publications/documents/dataskillscareersfinalreport.aspx> (Archived by WebCite® at <http://www.webcitation.org/6sw26BBkS>)
- Tech Partnership on behalf of SAS UK and Ireland (2014). [Big data analytics assessment of demand for labour and skills 2013-2020](#). London: Tech Partnership. Retrieved from [https://www.thetechpartnership.com/globalassets/pdfs/research-2014/bigdata\\_report\\_nov14.pdf](https://www.thetechpartnership.com/globalassets/pdfs/research-2014/bigdata_report_nov14.pdf) (Archived by WebCite® at <http://www.webcitation.org/6sw2EKJhK>)
- Witt, M. (2012). Co-designing, co-developing, and co-implementing an institutional data repository service. *Journal of Library Administration*, 52(2), 172-188.

---

## How to cite this paper

Kennan, M. A. (2017). 'In the eye of the beholder': knowledge and

skills requirements for data professionals In *Proceedings of RAILS - Research Applications, Information and Library Studies, 2016, School of Information Management, Victoria University of Wellington, New Zealand, 6-8 December, 2016.. Information Research, 22(4)*, paper rails1601. Retrieved from <http://InformationR.net/ir/22-4/rails/rails1601.html> (Archived by WebCite® at <http://www.webcitation.org/6vOCCFdnr>)

## Appendix 1: Interview and focus group questions

Gartner forecast that by 2015, 4.4 million professionals with experience in data management and analysis will be needed worldwide. Some propose that the data professionals, data managers or data scientists required to work in this area are a “new breed” for whom the knowledge and skill requirements are just emerging. As an academic in an institution which educates librarians and information managers, we are interested in whether we can put together courses to address this demand. To do so we need to understand what the knowledge and skills requirements in different types of organisations are for data professionals and this is why we are asking for your time for this interview.

1. Can you please briefly describe your organisation, your role and the kinds of data which are managed and used as a part of your role?
2. Are there particular other people who work with the same kinds of data you do, and what are their roles/role titles?
3. Currently, what kinds of education and training do the people who fill these data centric roles have?
4. What are the key knowledge and skills in your opinion required by professionals working with data in similar roles?
5. Generally, are all these data skills required in the one role or are there specialist roles? What are these roles?
6. How much information about data do people working in non-data centric roles need to have?
7. Do people come to your organisation equipped for the data roles you have?
8. If not, do you train them in-house or use off-site training?

Interviewer: There is some debate in the educational community about what type of educational background would be idea for data professionals. Ideas mooted have been as diverse a computer science, information management, librarianship and archival education, business and information systems, digital curation, a specific data course focussing on all aspects from creation to analysis to curation and preservation.

9. Ideally, for the data roles in your organisation – starting with your own role - what kinds of education and/or qualifications would graduates who take these roles have?
10. What specific subjects or areas of knowledge are key?
11. What are the continuing education/continuing professional development needs of people in data roles in your organisation?
12. What are the typical career paths for data professionals within your organisation?

## Appendix 2: Interviewees

	Role	Organization
1	Data Specialist– research data management capacity building role	Eresearch Support Organisation (individual interview)
2	Scientific Data Coordinator	Scientific Research Organisation 1 (individual interview)
3	Project Officer	Scientific Research Organisation 2 (individual interview)
4	Head of Resources Division, Information Management team	Scientific Research Organisation 3 (group interview)
5	Researcher and Data Specialist	Scientific Research Organisation 3 and University (group interview)
6	Director Science Data Platforms and Strategy	Scientific Research Organisation 3 (group interview)
7	Senior Data Strategist/Data Scientist	Scientific Research Organisation 3 (group interview)
8	Manager Information Services	Scientific Research Organisation 4 (group interview)
9	Data Librarian	Scientific Research Organisation 4 (group interview)
10	Data Librarian	Scientific Research Organisation 4 (group interview)
11	Information Specialist	Scientific Research Organisation 4 (group interview)
12	Director Digital Library Services	University Library 1 (individual interview)

13	Project Officer Library Repository Service	University Library 1 (group interview)
14	Technical Lead, Library Repository Service	University Library 1 (group interview)
15	Manager Data Quality	University Library 1 (group interview)
16	Library Repository Manager	University Library 1 (group interview)
17	Manager, Research Reporting	University Library 1 (group interview)
18	Manager, Content and Discovery (previously Data Librarian)	University Library 2 (individual interview)
19	Archivist	University Library 3 &IT service (group interview)
20	eResearch Manager	University Library 3 &IT service (group interview)
21	eResearch Analyst	University Library 3 &IT service (group interview)
22	eResearch Contractor	University Library 3 &IT service (group interview)
23	Information Service Librarian	University Library 3 &IT service (group interview)
24	Data Librarian	University Library 3 &IT service (individual interview)
25	Data Librarian	University Library 4 (previous role) and ANDS (current role)(Individual interview)
26	Data Architect	Bank 1 (individual interview)
27	Business Analysis Manager	Bank 1 (individual interview)
28	Workload Performance Engineer	Bank 2 (group interview)
29	Programmer	Bank 2 (group interview)
30	Principal Analyst	Utility company (individual interview)
31	Manager Digital Archives	State Records authority (individual interview)
32	Manager, Knowledge and Information management	Government authority (group interview)
33	Assistant Director General	National Archive (group interview)
34	Director, Information	Government Department (group interview)
	Director, Information	Government Department

35	Management	(group interview)
36	Big data statistical researcher, health sciences	University (individual interview)

[Find other papers on this subject](#)

Check for citations, [using Google Scholar](#)

---

© the author, 2017.

**131** Last updated: 1st December, 2017

---

[Contents](#) | [Author index](#) | [Subject index](#) |  
[Search](#) | [Home](#)

---