

Library capacity for data curation services: a US national survey

Library
capacity for
data curation
services

Ayoung Yoon

*Department of Library and Information Science,
Indiana University – Purdue University at Indianapolis, Indianapolis,
Indiana, USA, and*

Devan Ray Donaldson

*Department of Information and Library Science,
Indiana University at Bloomington, Bloomington, Indiana, USA*

Received 31 December 2018
Revised 20 March 2019
Accepted 21 March 2019

Abstract

Purpose – The purpose of this paper is to understand the landscape of data curation services among public and academic libraries in the USA, with a focus on library capacity for providing data curation services.

Design/methodology/approach – The authors conducted an online survey by employing stratified sampling from the American Library Directory. A total of 198 responses were analyzed.

Findings – The authors' findings provide insight into the current landscape of libraries' data curation services. The survey participants evaluated six capacity dimensions for both public and academic libraries – value, financial, administrative, technical infrastructure, human resources and network. The ratings the participants gave to these capacity dimensions were significantly different between academic and public libraries.

Practical implications – This study suggests several areas in which libraries will benefit from further developing their capacity to successfully run data curation services.

Originality/value – This is among the first research study to address the concept of capacity in the context of libraries' data curation services.

Keywords Library services, Organizational capacity, Data curation, Data services, Library capacity, National survey

Paper type Research paper

Introduction

The interest in, and need for, data services in libraries is increasing rapidly. Consequently, libraries – mostly academic, but also a few public – are attempting to establish these services. Motivations behind the services vary due to the different missions of each type of library: academic libraries tend to serve faculty and students, and are significantly influenced by the funding agencies' mandate for research data sharing; while public libraries are engaged with open data initiatives to serve their community and citizens, and are mostly influenced by the open government movement. Several recent studies have identified an increase in academic libraries launching data services or extending existing services (Cox and Pinfield, 2014; Yoon and Schultz, 2017; Xia and Wang, 2014), which are (or should be) accompanied by proper data curation programs. Although it is relatively new in the public library domain, recent studies have argued the need for public libraries' active engagement in providing data services to communities (Bertot *et al.*, 2014; Gibson *et al.*, 2009; Robinson and Mather, 2017). These studies also demonstrate a significant level of variation in these services and programs, perhaps due to the libraries' differing organizational capacities (Yoon and Schultz, 2017).

To successfully and sustainably launch and maintain data services in libraries, data curation programs must be tailored to a library's existing capacity. Noah and Brickman (2004)



argue that capacity building is necessary to strengthen or support a library's ability to provide high-quality, accessible and sustainable services to its community – a point that is also applicable to data curation services. Eventually, libraries must appropriately build or grow their capacities to sustain or extend their data curation programs in the long-term. Doing so will ensure effectiveness and help libraries avoid unintended consequences.

Despite the importance of capacity to libraries, few scholars have discussed what the concept of capacity actually means in the library context, why it is important, and how capacity assessment and building can be accomplished (Carrigan, 2015). Further, no study has analyzed how capacity relates to data curation services in libraries. Therefore, the aim of this study is to understand the landscape of data curation services among public and academic libraries in the USA, focusing on library capacity for providing data curation services. While previous literature has discussed organizational capacity as critical to fulfilling organizational purpose, this concept has not been extensively employed when studying libraries (Noah and Brickman, 2004).

In this study, we broadly define data curation as the active and ongoing management of data through its lifecycle to best serve the needs of current and future users. Based on our review of the literature on capacity and libraries, we define capacity as a library's ability to implement or perform internally – or externally-generated expectations and outcomes (Barman and MacIndoe, 2012; European Commission, 2005; Horton *et al.*, 2003; Sharpe, 2006). As understanding capacity is a fundamental prerequisite for effectively realizing expected outcomes, this study's findings help build a foundation to expand libraries' capacity for data curation services.

As noted, our study included both academic and public libraries. Because we understand the differences between academic and public libraries in terms of workforce, priorities and funding, one of our goals is to use this study to examine these differences. This will guide our next study in developing a capacity assessment framework for libraries, that is whether one framework is applicable to both types of libraries, or two separate frameworks are necessary for academic and public libraries.

Literature review

Academic libraries have been traditionally responsible for curating data to address their mission of protecting and disseminating data, particularly with recent mandates from funding agencies' data sharing policies. Recent research suggests the way to implement data services to meet the new demand is assess researchers' needs, select a repository environment, prepare training materials both for librarians and researchers, build expertise among library staff and establish metadata guidelines (Erway *et al.*, 2016). The need to provide data services, however, is not only for academic and research libraries that provide data management services for faculty. Given the rise in publicly available data resources, many open government data initiatives at the federal, state, and city levels currently exist or are in the process of being created, which involve partnerships with public libraries. For example, the Boston Public Library has a recent initiative to develop an open data collection for the City of Boston (City of Boston, 2017). Although the involvement of public libraries in community data services is in its infancy, more public libraries are developing their interests in data service models. Further, several studies underscore the significance of public libraries to be Open Data Hubs. Public libraries have the potential to open doors to marginalized populations and encourage citizen to contribute to existing data sets, improving the quality of life for everyone (Greenwalt, 2014; Sey *et al.*, 2013). Despite the continuous efforts and new demands in academic and public libraries, previous studies also point out the gaps and opportunities for libraries to hone their data curation services to more effectively meet users' needs (Johnston *et al.*, 2018). Further, existing efforts to understand libraries' data curation practices focus on

academic libraries (Primary Research Group, 2018), leaving public libraries understudied, mainly because their practices are in their infancy.

The concept of organizational capacity has received considerable attention during the past several decades, especially in business management, organizational studies and international developmental studies. Methods for assessing organizational capacity have been developed, and some frameworks for building organizational capacity have been proposed (De Vita and Fleming, 2001; McKinsey and Company, 2001; Nu'Man *et al.*, 2007). While several studies have proposed slightly different definitions of organizational capacity, common among them is the emphasis on an organization's ability to implement or perform internally – or externally-generated expectations and outcomes from their organizational fields (Barman and MacIndoe, 2012; European Commission, 2005; Horton *et al.*, 2003; Sharpe, 2006). This suggests that the conscious development of organizational capacity is critical to enabling an organization to fulfill its mission. Previous researchers have emphasized that assessment is essential to building organizational capacity, arguing that assessing capacity is a prerequisite for the interlinked decisions made in any organization – such as strategic and operational choices, ongoing policy dialogue and further capacity development processes (European Commission, 2005).

While existing literature has proposed different dimensions of organizational capacity using different names and attributes, it is still possible to identify some common dimensions across literature from various disciplines. The first common dimension is human resource capacity. Generally referring to the ability to deploy human capital within the organization, this dimension also includes not only the competencies, skills and knowledge, but also the leadership, attitudes, motivation and behaviors of individuals in the organization (Arnold *et al.*, 2014; Bourgeois and Cousins, 2013; Hall *et al.*, 2003; Misener and Doherty, 2009; Paynter and Berner, 2014). The second dimension is financial capacity, referring to an organization's ability to develop and deploy financial capital, including issues such as the revenues, expenses, assets and liabilities of the organization (Hall *et al.*, 2003; Misener and Doherty, 2009). Many have argued the importance of financial capacity, as it is directly related to completing planned activities (e.g. Bourgeois and Cousins, 2013; Paynter and Berner, 2014). Third, several researchers agree on the importance of administrative capacity. Often also called management and planning capacity, administrative capacity can be understood as the ability to develop and perform strategic plans, program plans and policies while mobilizing other resources (e.g. financial and human resources) to support daily management and organizational missions (Balduck *et al.*, 2015; Hall *et al.*, 2003; Misener and Doherty, 2009; Paynter and Berner, 2014). The fourth dimension common in the literature on this topic is infrastructure or resource capacity – an organization's ability to deploy or rely on organizational elements related to everyday operations, including physical and technical infrastructure, policies and manuals (Fredericksen and London, 2000; Hall *et al.*, 2003; Misener and Doherty, 2009; Paynter and Berner, 2014). Often this dimension is related to the stability of the governance structure (Bourgeois and Cousins, 2013). The last capacity dimension is network, also called relationship or partnership capacity. It is the ability to draw on relationships with relevant stakeholders – including clients, members, funding agencies, partners, government, media, corporations and the public. Many have argued that it is important to manage operations for maximizing relationships with key stakeholder groups because this enables access to shared resources, knowledge and experience (Balduck *et al.*, 2015; Balser and McCluskey, 2005; Hall *et al.*, 2003; Misener and Doherty, 2009; Paynter and Berner, 2014). Though considered separately in this literature review, these dimensions are not mutually exclusive, as they all influence each other and are interrelated in supporting organizational functions and missions.

As the concept of organizational capacity has been developed and applied primarily in the context of nonprofit organizations, it is applicable to libraries, whether academic or public. Despite its usefulness and significance, capacity building or assessment has not

garnered much attention in the library field (Carrigan, 2015), though there have been efforts to build library capacity in certain aspects (e.g. building human resource capacity, library network or policy) (Denison and Hardy, 2002; Ogunmodede and Mafelu, 2012; Sarmah, 2014). Noah and Brickman (2004) argue for the importance of capacity building in libraries and have provided a framework to assess library capacity. Their framework includes ten dimensions, many of which can be mapped alongside the five capacity dimensions described above: clarity and alignment (e.g. having a clear vision and a shared mission), human resources, facilities (e.g. adequate equipment and technology), collections, service delivery, collaborations, management systems, communication, continuous improvement and flexibility and funding. The similarity of these ten dimensions with the abovementioned five suggest that the concept of capacity can be used to assess conditions necessary for a library to perform curation activities and realize its potential for success.

Libraries – particularly academic libraries – have a longstanding history of launching initiatives to conduct either formal or informal assessments to understand their services based on outcomes and performances. For example, in May 2017, SPEC Surveys, which gather information on current research library practices and policies from Association of Research Libraries (ARL) member institutions to help them compare practices and improve performance, published SPEC Kit 354: data Curation to explore the infrastructure that ARL member institutions are using for active and ongoing data management. While SPEC Kit 354 underscores the current issues that librarians must address regarding data curation, it does not consider the libraries' ability to perform tasks and produce outputs, to understand and solve problems and to make informed choices related to data curation, all of which are cornerstones of capacity. Our study builds upon SPEC Kit 354 by expanding the scope of libraries beyond ARL member institutions and addressing different capacity dimensions for libraries' data curation services.

Methods

To collect and analyze data on libraries' capacity for data curation, including their current practices and challenges, we developed and administered an online survey. The survey consisted of the following sections: institutional profile, curation practices and challenges and capacity assessment. Questions related to library capacity were designed based on our literature review, addressing six different capacity dimensions identified from the literature with their definitions modified to suit our study context. Each capacity dimension was defined as follows:

- Value dimension – libraries' ability to develop a shared vision, goal and mission on data curation services, adopting changes to revise them when necessary.
- Administrative dimension – libraries' ability to develop and draw on their daily management, workflow, strategic plans, and policies to support data curation services.
- Human resource dimension –libraries' ability to deploy human capital. This dimension not only includes librarians' competencies, skills and knowledge related to data curation, but also librarians' attitudes, motivation and behaviors that influence both the organizational culture and the data curation services.
- Financial dimension – libraries' ability to deploy financial capital for sustainably supporting data curation services, including revenues, expenses and assets.
- Technical infrastructure dimension – libraries' ability to deploy or rely on their technical resources related to data curation service operations.
- Network dimension– libraries' ability to draw on relationships with relevant stakeholders (both internal and external), such as serving populations, internal organizational units, funding agencies and professional communities.

We asked three to eight questions per dimension, depending on the resources and sub-constructs associated with each capacity dimension. The survey included Likert-scale, categorical and open-ended questions. We used skip logic in the survey so that participants would only answer relevant sections. For instance, libraries without data curation services only answered questions about their future plans and did not answer any questions about curation practices and capacity assessment. To ensure validity, the survey protocol was reviewed by an expert panel of professionals with experience and expertise in capacity, libraries and data curation programs and services.

Our sample was drawn from the “American Library Directory: 2016–2017” (American Library Directory, 2016–2017, 2016), which lists 16,878 public libraries and 3,635 academic libraries. We randomly selected libraries from these lists. Since academic libraries and public libraries have vastly different workforces, priorities and funding streams, we employed stratified sampling for our study. We created strata for public libraries according to categories specified in the “American Library Directory”: public libraries, excluding branches ($n = 9,669$), main public libraries that have branches ($n = 1,409$) and public library branches ($n = 7,209$). Likewise, we created strata for academic libraries according to categories specified in the “American Library Directory”: community college libraries ($n = 1,115$) and university and college libraries ($n = 2,520$). A total of 1,652 libraries were identified through this process. After cleaning the list and removing libraries without websites or e-mail addresses, our final list included 1,452 libraries – 706 public libraries and 746 academic libraries. To identify the best contacts for the survey, we examined the contact information provided in the “American Library Directory” and conducted additional searches to determine who was responsible for data curation at these institutions.

The survey was distributed online using Qualtrics, an online survey software tool (www.qualtrics.com). All survey invitations were personalized and sent to either deans/directors of libraries or data librarians. When invitations were sent to deans or directors, we asked them to forward the invitation to their staff who handled curation activities. We also distributed the survey through relevant library association LISTSERVs, Research Data Access, Preservation Summit and the Public Library Association. The survey was open from June 2018 to August 2018, with two reminders during the course of the survey period. Of the 1,452 invitations sent, 40 e-mails bounced back. In total, 1,412 invitations were successfully sent.

The collected data were analyzed quantitatively and qualitatively. The data were exported in CSV format and imported to SPSS for descriptive analysis, using univariate and bivariate descriptive statistics (e.g. frequency distribution). A qualitative analysis of the responses to the open-ended questions was inductively analyzed using Microsoft Excel, via searches for emerging themes from the text responses.

Results

Our survey collected a total of 230 responses, bringing our response rate to 16 percent. After review of the data, we dropped 32 responses from further analysis because they were incomplete and lacked research value; a total of 198 responses were analyzed. College and university libraries made up the majority of respondents (114 responses, 58 percent) with public libraries representing 40 percent (80 responses). Four respondents identified themselves as belonging to the “other” category. Among these 198 libraries, 35 percent (70 libraries; 24 public libraries and 46 academic libraries) reported that they currently had a data curation service, while 15 percent (29 libraries) were in the process of developing one, and 50 percent reported not having one. With the exception of a few questions (e.g. service years, motivations, target audiences, types of data supported, etc.) that included the responses of staff at 99 libraries with data curation services existing or in development, this paper reports specifically on responses of staff from those 70 libraries that already had data curation services in place.

Data curation services

Among the 74 libraries that answered the question about the number of years they had offered data services, most began providing data curation services recently, between 2011 and 2015 (24 libraries, 32 percent) or after 2016 (13 libraries, 18 percent). Overall, the target audiences for the libraries' data curation services were mostly academic researchers, whether affiliated with their institution or not (67 percent), followed by the general public (20 percent). While academic researchers and other university affiliates (e.g. students) were the major audience for academic libraries, the general public was the major audience for public libraries. Government organizations and nonprofit organizations also appeared to be the public libraries' audiences. Other academic library audiences (9 percent) included students, instructors and staff. Figure 1 presents the details of the target audiences.

This difference in target audiences between academic and public libraries was well reflected in their motivations for launching data services. As data from answers to our open-ended questions suggest, the data curation services of most academic libraries were launched to support faculty and students to comply with funding agencies' mandates and their data management requirements, to store large data sets and to either pursue or maintain R1 status. On the other hand, public libraries aimed to preserve local history, to serve the public, to expand information discovery services and to keep the information they were providing to the public pertinent and up-to-date.

The majority of libraries' data curation services were intended to support research driven by the academic libraries, as the majority of data types the libraries supported were research data from different disciplines, such as the life sciences, social sciences, arts and humanities and engineering and applied sciences. However, libraries also curated some other types of data, such as government data, administrative data and some business data (see Figure 2). Not surprisingly, academic libraries provided the most support for research data, although some academic libraries stated that they provided support for other types of data, such as government data and social media data. Public libraries provided more support for administrative, business, and arts and humanities research data.

When we asked about the data curation activities that the libraries currently supported, we found that many curation services focused on data management planning (14 percent) and data organization (e.g. metadata, 13 percent), as well as data sharing (12 percent) and

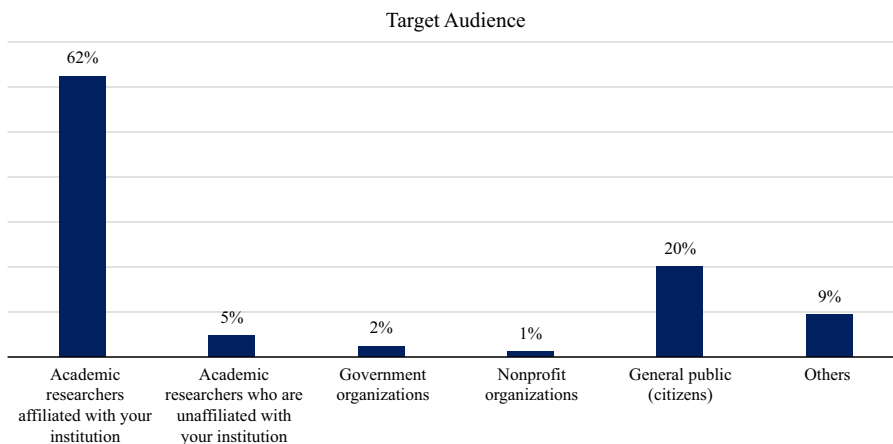


Figure 1.
Who are the target audiences of the data curation services? (check all that apply)

Notes: Responses eligible for analysis, $n=71$; total response, $n=85$

Library capacity for data curation services

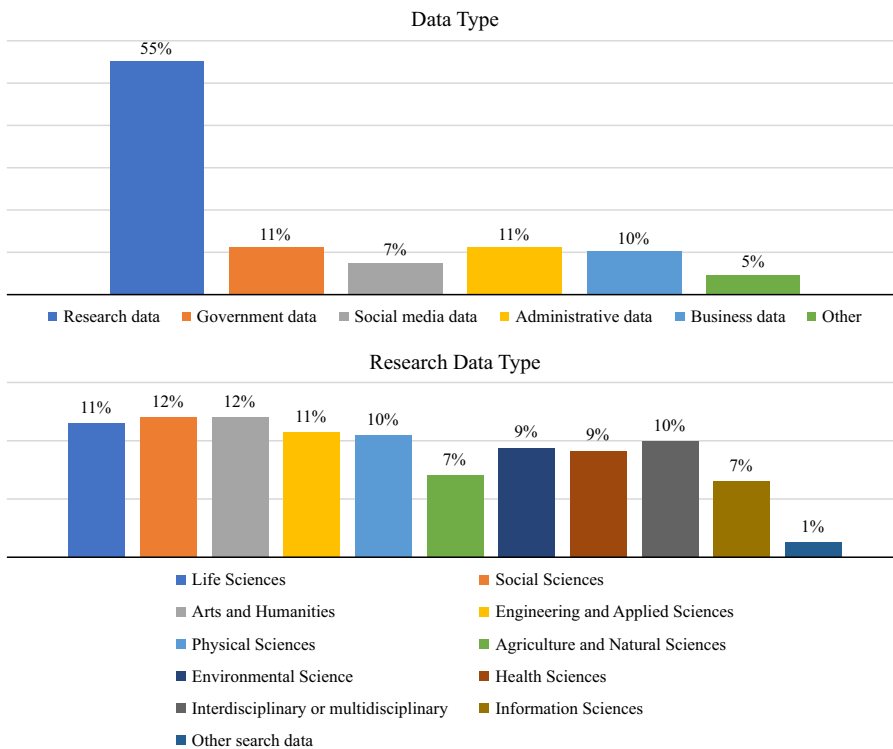


Figure 2. What types of data does your library support through its services? (check all that apply)

Notes: Responses eligible for analysis, $n=64$; total response, $n=431$

reuse (12 percent). See Figure 3 for more information on the curation services that the participants reported their libraries supported.

Among the 35 libraries that stated they provided data repository support, 71 percent housed data in their data repositories including institutional repositories or internal servers.

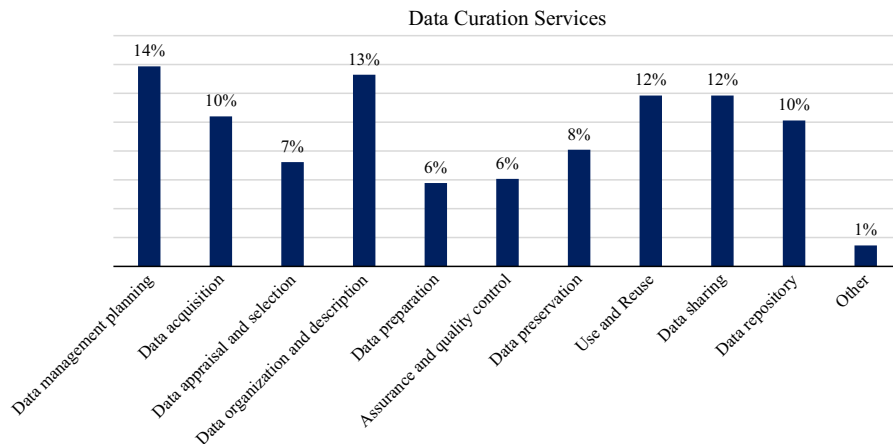


Figure 3. Please check all areas of data curation that your library currently supports

Notes: Responses eligible for analysis, $n=60$; total responses, $n=346$

No public libraries supported repository services, although three public libraries stated they housed data on their internal servers. According to the participants, 28 libraries used external hosting services – such as Amazon Cloud, Digital Commons, FigShare and Open Science Framework. Most of them were academic libraries, with only four public libraries using external hosting services.

Library capacity for data curation services

Participants varied in how they rated their libraries across the six capacity dimensions. Regarding the value dimension, a majority of participants from public and academic libraries either agreed or strongly agreed that their libraries were committed to data curation services (83 percent) and that they have aligned those services with their missions (88 percent). However, more than half of the participants (58 percent) either agreed or strongly agreed that their libraries had a shared belief on the value of data curation services (see Figure 4).

Participants' evaluation of the human resource capacity dimension varied widely. While the majority of participants thought that their librarians were well equipped with the necessary skills and knowledge to lead data curation services (76 percent), almost half of the participants reported that they do not have enough staff to provide or maintain data curation services (49 percent). This influenced the overall assessment of their libraries' human resources capacity, since 56 percent of participants felt that they do not have sufficient human resources to run data curation services (see Figure 5).

When we asked about the number of staff members whose work responsibilities focus exclusively (100 percent) on data curation services, almost half of the libraries (53 percent) had zero staff members, 25 percent had one staff member, 11 percent had two staff members and 12 percent had more than three staff members. Academic libraries generally had more staff than public libraries, as nearly 52 percent of academic libraries who answered this question had at least one staff member who worked full-time in data curation services. Only about 22 percent of public libraries had at least one full-time data curation staff member. The difference between academic and public libraries was more evident based on participants' answers to questions about whether their libraries had any new position openings related to data curation in the past six months. In total, 30 percent of academic library participants reported they had new openings; in contrast, none of the public library participants reported having any new data curation openings. The main motivation for having a new position was the opening of an existing position due to staff turnover

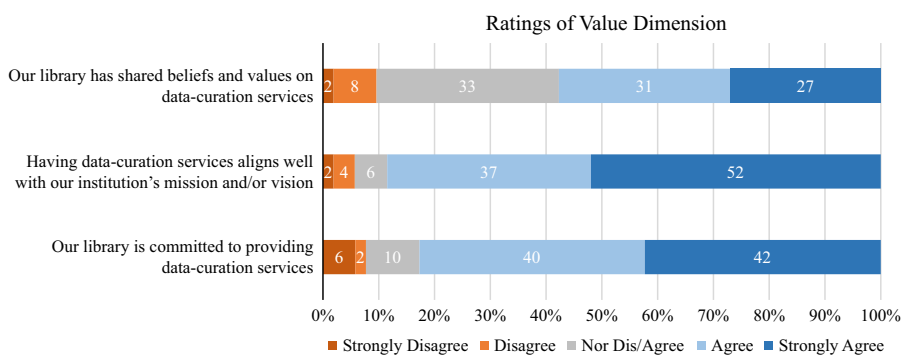


Figure 4.
Capacity ratings on value dimension

Note: $n=52$

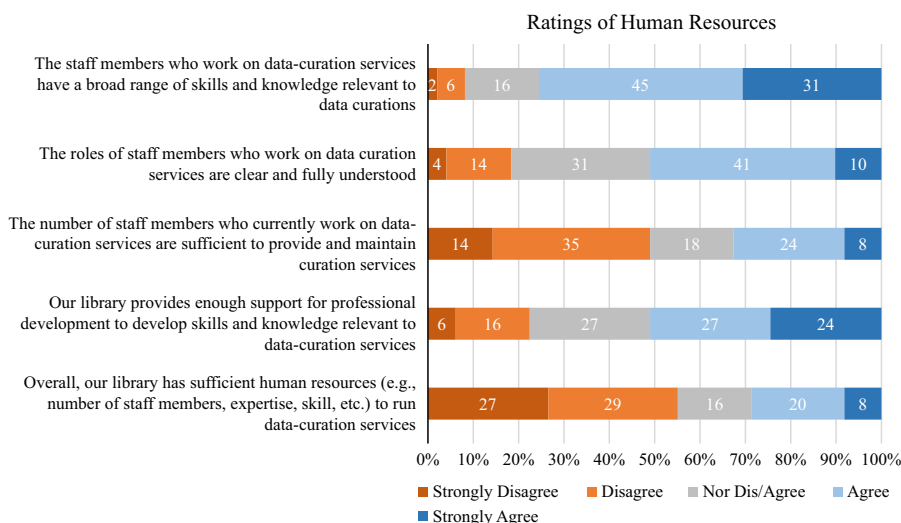


Figure 5.
Capacity ratings on
human resources
dimension

Note: $n = 49$

(31 percent), followed by repurposing of an existing position for data curation (19 percent), while only 8 percent cited adding a position.

Regarding the administrative dimension, participants generally agreed that their libraries had a clear direction for providing and further developing data curation services (49 percent), as well as strong leadership (46 percent) supported by a strategic plan (54 percent). However, most of the participants reported that their public and academic libraries did not have concrete and detailed workflows for providing data curation services (43 percent), nor did they have a plan for performance measures to define the success of data curation services (62 percent). Having a plan to keep up with policy changes was another area in which both library types were not very strong. More public library participants than academic library participants reported not having a plan to work with policy changes (see Figure 6).

Overall, our results suggest that the financial dimension was the weakest area in the participating libraries' assessments. In total, 50 percent of the participants either disagreed or strongly disagreed that their libraries had a solid financial plan to support infrastructure for data curation services, and 55 percent either disagreed or strongly disagreed that they had designated funding for long-term data management. Funding for supporting staff was a lesser concern, as only 33 percent disagreed that they had a solid financial plan for staff. Participants from public and academic libraries followed the same pattern of agreement and disagreement with these statements (see Figure 7).

When we asked participants to briefly describe their libraries' business model for data curation services, most participants – whether public or academic – said that their business model was “no charge.” However, a few academic university library participants said that their business model was based on the size of the data. For example, data over a certain size (e.g. 1TB) incurred a fee; otherwise there was no charge.

About half of the library participants had a strong technical infrastructure for data curation services (44 percent), as well as technology management staff to support technological needs (54 percent). Despite the overall agreement on strong technological infrastructure and sufficient staff support, only 35 percent agreed that they had sufficient

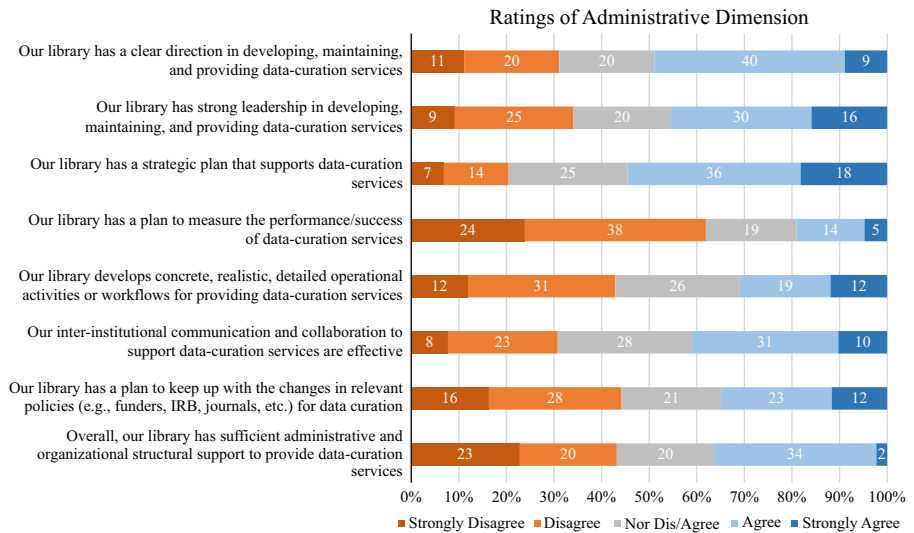


Figure 6.
Capacity ratings
on administrative
dimension

Note: $n=45$

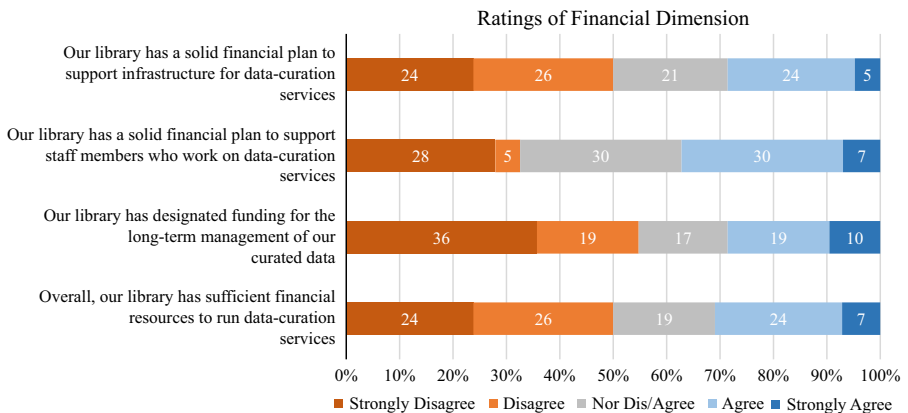


Figure 7.
Capacity ratings on
financial dimension

Note: $n=43$

technological resources. However, this rating was mostly driven by academic library participants because the majority of public library participants either disagreed or strongly disagreed with all three statements. The technical infrastructure dimension showed the most significant difference between academic and public libraries (Figure 8).

Regarding the network dimension, more than half of the participants reported that their libraries had built necessary internal partnerships to support data curation services (55 percent), as well as external partnerships (50 percent). In total, 54 percent of participants said that their libraries did some marketing to the intended audiences. However, many participants also answered that their libraries did not have a strong presence among their user communities (54 percent), nor among professional communities related to data curation (54 percent). Interestingly, while most public libraries disagreed with these statements, they

Library capacity for data curation services

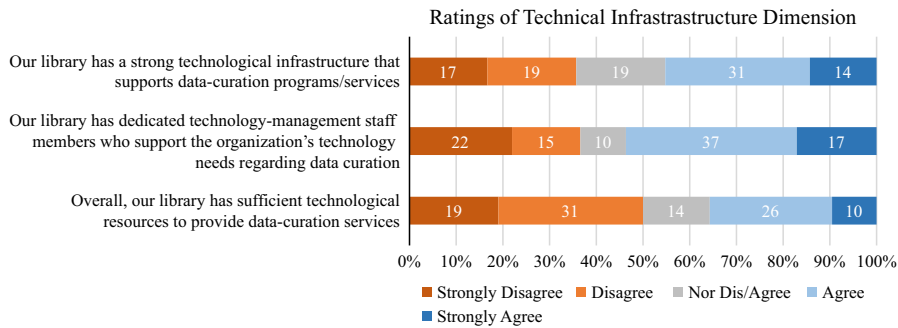


Figure 8. Capacity ratings on technical infrastructure dimension

Note: $n=42$

mostly agreed that they had built external partnerships with various entities for data curation services (Figure 9).

Challenges

When we asked about perceived challenges to developing, maintaining and providing data curation services – requesting that participants rate items on a 1–5 scale (e.g. from 1 = not at all challenging to 5 = very challenging) – participants rated adoption of data curation services, diversity of needs and perception of data curation across disciplines, and building a shared vision of data curation as the three most challenging aspects for their libraries. The public library participants rated the following items as some of the most challenging issues:

- diversity of needs and perceptions across disciplines;
- securing funding for data curation services;
- technical support or infrastructure at the institutional level; and
- outreach and engagement.

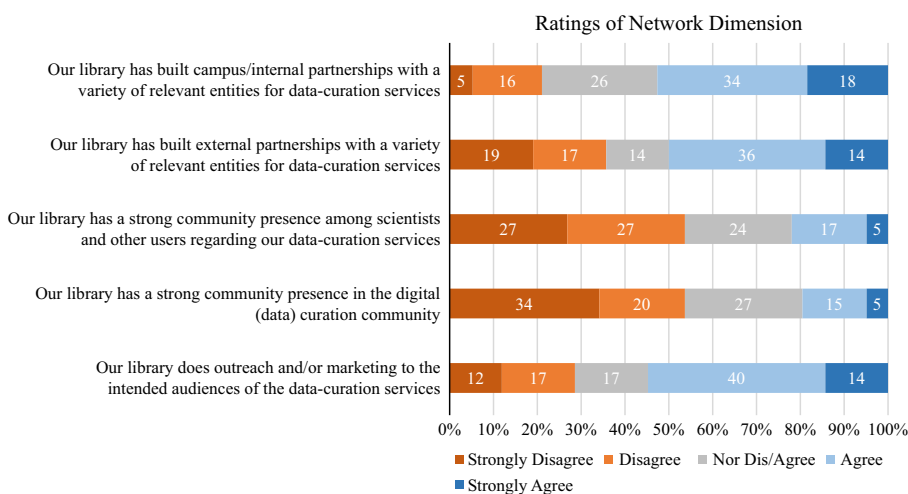


Figure 9. Capacity ratings on network dimension

Note: $n=42$

When asked about perceived challenges to developing, maintaining and providing data curation services, college and university library participants rated the following as some of the most challenging issues:

- building a shared vision and values on data curation;
- adoption of data curation services; and
- diversity of needs and perceptions across disciplines.

Some of the challenges listed in Figure 10 were further reinforced by participants' comments in our open-ended questions. For instance, some academic library participants discussed the challenges of obtaining upper-level administrative help because of the relationship with other campus entities, such as the Office of the Vice President for Research. Even within libraries, generational divides between librarians who handle data and those who "don't do" data produced some tension, which made obtaining full support difficult, despite campuses being engaged in research data services. Employee ambivalence (e.g. "What is the value of doing data curation?") was mentioned, not only in relation to administrative support, but also regarding securing financial support.

One interesting challenge discussed among academic library participants was the notion of scale and capacity. One librarian said, "We'd do full curation services for no charge if only we had enough people to help. Now it's all about training and empowering researchers to self-serve their curation. We have that capacity and are expanding ours through training and outreach." This statement well reflected the factors that drive their current library curation service model.

While many academic library participants elaborated on the challenges they face, public library participants reported that many of them either had not actually begun providing data curation services or that their services were too limited to report any challenges because their libraries were small.

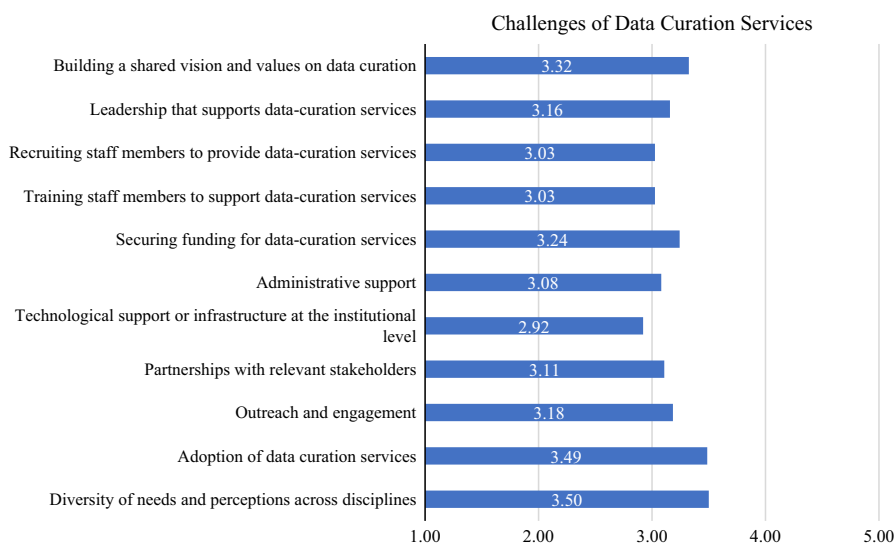


Figure 10.
Mean value of
challenges

Notes: $n = 39$, 1 = not at all challenging, 5 = very challenging

Discussion

Our findings provide insight into the current landscape of library data curation services. While only 35 percent of libraries in our survey stated that they have data curation services, the numbers are not discouraging, since our results indicate that more libraries (15 percent) are currently in the process of developing data curation services. Since the majority of data curation services were developed after 2011, it is likely that there will be more development in the future. In addition, although mostly academic libraries (63 percent) offered data curation services (since they have been expected to work with research data to meet funders' data management requirements), public libraries seemed to be aware of this new service area and showed interest.

Currently, academic libraries are leading the development of data curation services. Active involvement of academic libraries in data curation was well reflected in the target audiences because 77 percent of target audiences for data curation services were researchers, including students. Accordingly, libraries provided more support for research data from various disciplines than for other types of data. However, despite the numbers being relatively small, libraries' curation services did also serve other audiences, such as the general public. These services were mostly driven by public libraries. It is not surprising to see such a difference between academic and public libraries, as they have distinctive missions and different communities to serve. The fact that public libraries serve citizens and provide support for other types of data – such as business, government and administrative data – demonstrates the potential role of public libraries in data curation, as there are growing expectations for public libraries to play a role in supporting communities' data work (City of Boston, 2017; Yoon *et al.*, 2018).

Our results demonstrate that the depth of data curation services and coverage provided varied among libraries, while libraries' curation services addressed the full curation lifecycle. Data management planning and data organization/description were two of the most commonly provided services. These results were driven by academic libraries' heavy involvement in research data management. It is interesting to note that data reuse support, with data sharing support, was the third most commonly provided curation service, as previous research reported that data reuse were one of the areas least supported by libraries (Yoon and Schultz, 2017). While this study was not able to further investigate what types of reuse services were available for users, it is an encouraging finding for library service development in this area.

Libraries' capacity for data curation services varied depending on each dimension. Among the six capacity dimensions – value, human resources, administrative, financial, technical infrastructure and network – it is worth mentioning that the value dimension was the one that participants reported their libraries having the most capacity to address, whether they were from academic or public libraries. However, building a shared vision and value on data curation was not always easy. Some librarians reported it as a challenge in our open-ended survey questions because, when building a vision and shared value, it often takes time to change people's perspectives and to build culture. Yet, most libraries were capable of building a vision for data curation that aligned well with their mission. Having a shared value and vision of data curation is the most important first step toward building successful data curation services, as it promotes investment in other dimensions, such as administrative support, financial support and support in human resources, which eventually helps to build capacity in all dimensions.

Unfortunately, the financial dimension was the weakest capacity for both academic and public libraries – a fact reflected in participants' rating of challenges, as they reported securing funding for data curation services as the third most challenging aspect of data curation services. Participants reported that their libraries needed more financial resources to support data curation infrastructure and staff. Securing designated funding sources for

long-term curation was another area that needed to be further strengthened to build capacity in the financial dimension.

In relation to limited financial capacity – especially for supporting staff designated to data curation services – participants stated that lacking staff was one of the major areas requiring further support to build human resource capacity. About half of the participants responded to the human resource questions, saying that their libraries did not have any staff whose responsibilities exclusively focused on data curation. This demonstrated the need to add more staff. This is truer for public libraries than academic libraries, as academic libraries have a longer history of offering data curation services in general due to their history of working with research data. Securing funding for data curation at public libraries can be more challenging. Regarding the role of public libraries in data curation, working with citizens and local communities has only recently been discussed. These libraries might not have users that they immediately need to serve. Recent literature suggests that there are growing demands from the public side (e.g. general citizens, community organizations) to utilize civic data for community development (Gurstein, 2011; The United Nations Secretary-General’s Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG), 2014; van Schalkwyk *et al.*, 2014). Public libraries should anticipate these future needs and could be more active players in supporting public demand for data through their curation services.

It is interesting to note that participants from both public and academic libraries considered their librarians as equipped with a broad range of skills and knowledge related to data curation. This is a positive sign for libraries’ data curation services and their ability to build their human resource capacity. Recruiting and training staff to provide data curation services was one of the least challenging areas identified by our participants. With growing attention to and acknowledgement of the importance of data curation, different levels and types of educational opportunities have been developed – graduate programs, certificates, MOOCs, workshops and online training modules (Keralis, 2012). These are important resources for supporting libraries in building their data curation skills and knowledge, which can influence libraries’ self-assessments regarding their capacity related to skills and knowledge development.

In the administrative capacity dimension, both public and academic libraries do not seem to have established and documented procedures for developing and providing data curation services. Participants from both types of libraries reported that they did not have a solid way to measure their data curation service performance or success, which was also partially related to the lack of a concrete, realistic and detailed workflow for the provision of data curation services. This is interesting, especially considering the long history of conducting either formal or informal assessment at libraries, particularly academic libraries, through various assessment methods (Hiller and Self, 2004). Perhaps the reason is because data curation services themselves are recent and, likely, libraries have not yet been able to fully develop a set of workflow and performance measurements. Performance measures and assessments are critical components that demonstrate the value of the services. As Kaufman and Watstein (2008) argue, “libraries cannot demonstrate institutional value to maximum effect until they define outcomes of institutional relevance and then measure the degree to which they attain them” (p. 227). Demonstrating the value of services will help to further build a shared vision within libraries, institutions and society at large, which should, in turn, support more investments. Thus, this finding suggests that libraries should put more effort into documenting outcomes and investigating performance related to their data curation services.

While participants in this survey agreed that their libraries had the appropriate capacity for technical infrastructure compared with other dimensions, technical infrastructure is the

dimension in which public and academic libraries differed the most. This gap between academic and public libraries is not surprising due to the different emphases on their existing service models, focal points and audiences. Given this, perhaps public libraries will require more technical resources to initiate or further develop their data curation services. Despite academic library participants rating their libraries as well equipped with technological infrastructure and technology staff support, they also reported a need for further resources to provide data curation services. What other sources they need should be further investigated.

Finally, in the network capacity dimension, participants reported that their libraries were generally weak in their community presence – both the user communities and the professional communities associated with data curation. However, they were capable of building internal and external networks for providing data curation services. Public libraries were particularly strong at building external partnerships for data curation services, perhaps because they were likely to rely on external services or infrastructure for doing data curation. Often building an external partnership is the most cost-effective decision when an organization is low on resources for conducting work. Further investigation of external partnerships is necessary to fully unpack the nature of these partnerships. Still, our findings suggest that libraries will need to further develop their community presence. The weak presence among the user and professional communities may be due to the recentness of data curation services, or perhaps libraries may not have much capacity to go out and build a strong presence when they are focusing on internal development, building and implementing a new service. Regardless, having a strong community presence is a key to successful service and would help libraries in various ways including, but not limited to, improved outreach and services for intended audiences, and the ability to keep up with other initiatives that can be used as models for their own services.

Conclusion

This study explored the landscape of data curation services among public and academic libraries in the USA, focusing on their staff's assessment of six different data curation capacity dimensions. The study's findings suggest that libraries' current capacities for providing data curation services are varied depending on the area, but there are several areas that libraries will benefit from further developing. This development will increase their capacity to successfully run data curation services. Our findings suggest that different approaches are necessary for academic and public libraries to build their capacity, as they differ in the audiences they serve, the types of data upon which they focus, their current data curation services status, and the resources they have available.

While this study is among the first to address the concept of capacity for data curation services in libraries, it also has some limitations. Although our survey response rate was acceptable, when we considered only the libraries that offered data curation services, our sample was lower than we had hoped. We also employed stratified sampling, thus, the results may not be fully generalizable. However, this could be a reflection of reality regarding the nature of data curation services and programs across the country, since it may be true that, as of yet, only certain portions of all national libraries have data curation services. In addition, our survey responses are heavily driven by academic libraries because academic libraries, in general, are more involved in data curation than public libraries. While our goal was to investigate both academic and public libraries, public libraries may not be fully reflected in our survey results, since the number of participants from public libraries was small compared with participants from academic libraries. Our study provided comparative views from our analysis, but separate investigations on each type of library would provide more in-depth insights.

References

- American Library Directory, 2016-2017 (2016), Information Today, Inc., New Providence, NJ.
- Arnold, K.E., Lynch, G., Huston, D., Wong, L., Jorn, L. and Olsen, C.W. (2014), "Building institutional capacities and competencies for systemic learning analytics initiatives", *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, ACM, New York, NY, pp. 257-260, available at: <https://doi.org/10.1145/2567574.2567593>
- Balduck, A.L., Lucidarme, S., Marlier, M. and Willem, A. (2015), "Organizational capacity and organizational ambition in nonprofit and voluntary sports clubs", *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, Vol. 26 No. 5, pp. 2023-2043, available at: <https://doi.org/10.1007/s11266-014-9502-x>
- Balsler, D. and McCluskey, J. (2005), "Managing stakeholder relationships and nonprofit organization effectiveness", *Nonprofit Management and Leadership*, Vol. 15 No. 3, pp. 295-315.
- Barman, E. and MacIndoe, H. (2012), "Institutional pressures and organizational capacity: the case of outcome measurement", *Sociological Forum*, Vol. 27 No. 1, pp. 70-93.
- Bertot, J.C., Butler, B.S. and Travis, D.M. (2014), "Local big data: the role of libraries in building community data infrastructures", *Proceedings of the 15th Annual International Conference on Digital Government Research, Aguascalientes*, pp. 17-23, available at: <https://doi.org/10.1145/2612733.2612762>
- Bourgeois, I. and Cousins, J.B. (2013), "Understanding dimensions of organizational evaluation capacity", *American Journal of Evaluation*, Vol. 34 No. 3, pp. 299-319, available at: <https://doi.org/10.1177/1098214013477235>
- Carrigan, D.P. (2015), "Organizational capacity and the public library (featuring an interview with Lexington (KY) public library executive director Ann Hammond)", June 16 available at: <http://publiclibrariesonline.org/2015/06/organizational-capacity-and-the-public-library-featuring-an-interview-with-lexington-ky-public-library-xia-executive-director-annhammond/> (accessed May 10, 2017).
- City of Boston (2017), "Open data to open knowledge", available at: www.boston.gov/innovation-and-technology/open-data-open-knowledge (accessed May 10, 2018).
- Cox, A.M. and Pinfield, S. (2014), "Research data management and libraries: current activities and future priorities", *Journal of Librarianship and Information Science*, Vol. 22 No. 4, pp. 299-316, available at: <https://doi.org/10.1177/0961000613492542>
- De Vita, C. and Fleming, C. (2001), *Building Capacity in Nonprofit Organizations*, The Urban Institute, Washington, DC.
- Denison, T. and Hardy, G. (2002), "Building the capacity of the public library network", White Room Electronic Publishing Pty Ltd, Centre for Community Networking Research, available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.616.9274&rep=rep1&type=pdf> (accessed May 10, 2018).
- Erway, R., Horton, L., Nurnberger, A., Otsuji, R. and Rushing, A. (2016), "Building blocks: laying the foundation for a research data management program", OCLC Research, Dublin, OH, available at: www.oclc.org/content/dam/research/publications/2016/oclcresearch-data-management-building-blocks-2016.pdf
- European Commission (2005), "Institutional assessment and capacity development: Why, What and How?", Office for Official Publications of the European Communities, Luxembourg.
- Fredericksen, P. and London, R. (2000), "Disconnect in the hollow state: the pivotal role of organizational capacity in community-based development organizations", *Public Administration Review*, Vol. 60 No. 3, pp. 230-239.
- Gibson, A., Bertot, J. and McClure, C.R. (2009), "Emerging role of public librarians as e-government providers", *42nd Hawaii International Conference on System Sciences, IEEE, Waikoloa, HI*, pp. 1-10, available at: <https://doi.org/10.1109/HICSS.2009.183>
- Greenwalt, R.T. (2014), "Revisiting open data practices", May 8, available at: <http://publiclibrariesonline.org/2014/05/revisiting-open-data-practices/> (accessed January 9, 2019).

- Gurstein, M.B. (2011), "Open data: empowering the empowered or effective data use for everyone?", *First Monday*, Vol. 16 No. 2, available at: <http://firstmonday.org/ojs/index.php/fm/article/view/3316>
- Hall, M., Andrukow, A., Barr, C., Brock, K., de Wit, M., Embuldeniya, D., Jolin, L., Lasby, D., Lévesque, B., Malinsky, E., Stowe, S. and Vaillancourt, Y. (2003), "The capacity to serve: a qualitative study of the challenges facing Canada's nonprofit and voluntary organizations", Canadian Centre for Philanthropy, Toronto, ON.
- Hiller, S. and Self, J. (2004), "From measurement to management: using data wisely for planning and decision-making", *Library Trends*, Vol. 52 No. 1, pp. 129-155.
- Horton, D., Alexaki, A., Bennett-Lartey, S., Noële Brice, K., Campilan, D., Carden, F. and Watts, J. (2003), "Evaluating capacity development: experiences from research and development organizations around the world", International Service for National Agricultural Research.
- Johnston, L.R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R. and Stewart, C. (2018), "How important is data curation? gaps and opportunities for academic libraries", *Journal of Librarianship and Scholarly Communication*, Vol. 6 No. 1, pp. 1-24, available at: <http://doi.org/10.7710/2162-3309.2198>
- Kaufman, P. and Watstein, S.B. (2008), "Library value (return on investment, ROI) and the challenge of placing a value on public services", *Reference Services Review*, Vol. 36 No. 3, pp. 226-231.
- Keralis, D.C.S. (2012), "Data curation education: a snapshot", in Jahnke, L., Asher, A. and Keralis, S.D.C. (Eds), *The Problem of DATA*, CLIR pub, Washington, DC, p. 154, available at: www.clir.org/pubs/reports/pub154/education/
- McKinsey and Company (2001), "Capacity assessment grid", available at: www.socialimpactexchange.org/sites/www.socialimpactexchange.org/files/McKinsey%20Self%20Assessment%20Grid.pdf (accessed May 8, 2015).
- Misener, K. and Doherty, A. (2009), "A case study of organizational capacity in nonprofit community sport", *Journal of Sport Management*, Vol. 23 No. 4, pp. 457-482.
- Noah, C. and Brickman, A. (2004), "Capacity building for libraries", *Public Libraries*, Vol. 43 No. 2, pp. 102-107.
- Nu'Man, J., King, W., Bhalakia, A. and Criss, S. (2007), "A framework for building organizational capacity integrating planning, monitoring, and evaluation", *Journal of Public Health Management Practice*, Vol. 13 No. 1, pp. S24-S32.
- Ogunmodede, T.A. and Mafelu, M.E. (2012), "Capacity building programmes for library staff in University of Ibadan and University of Lagos libraries", *Samaru Journal of Information Studies*, Vol. 12 Nos 1-2, pp. 61-67.
- Paynter, S. and Berner, M. (2014), "Organizational capacity of nonprofit social service agencies", *Journal of Health & Human Services Administration*, Vol. 37 No. 1, pp. 111-145.
- Primary Research Group (2018), *International Survey of Academic Library Data Curation Practices*, 2019 ed., ResearchAndMarkets, New York, NY, available at: www.researchandmarkets.com/research/66r8np/_international?w=4
- Robinson, P. and Mather, L.W. (2017), "Open data community maturity: libraries as civic infomediaries", *Journal of the Urban and Regional Information Systems Association*, Vol. 28 No. 1, pp. 31-38.
- Sarmah, R. (2014), "Capacity building in university library service", presented at the 9th Convention PLANNER, Dibrugarh University, Assam, available at: <http://ir.inflibnet.ac.in:8080/ir/bitstream/1944/1826/1/22.pdf>
- Sey, A., Coward, C., Rothschild, C., Clark, M. and Koepke, L. (2013), "Public libraries connecting people for development: findings from the Global Impact Study", Technology & Social Change Group, University of Washington Information School, Seattle, WA.
- Sharpe, E.K. (2006), "Resources at the grassroots of recreation: organizational capacity and quality of experience in a community sport organization", *Leisure Sciences*, Vol. 28 No. 4, pp. 385-401, available at: <https://doi.org/10.1080/01490400600745894>

-
- The United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) (2014), "A world that counts, mobilizing the data revolution", available at: www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf (accessed May 10, 2018).
- van Schalkwyk, F., Wilmers, M. and Czerniewicz, L. (2014), "Case study: open data in the governance of South African higher education", available at: www.opendataresearch.org/sites/default/files/publications/Open%20Data%20in%20the%20Goernance%20of%20South%20African%20Higher%20Education%20WEB.PDF (accessed January 16, 2019).
- Xia, J. and Wang, M. (2014), "Competencies and responsibilities of social science data librarians: an analysis of job descriptions", *College and Research Libraries*, Vol. 75 No. 3, pp. 362-388.
- Yoon, A. and Schultz, T. (2017), "Research data management services in academic libraries in the US: a content analysis of libraries' websites", *College & Research Libraries*, Vol. 78 No. 7, pp. 920-933, available at: <https://doi.org/10.5860/crl.78.7.920>
- Yoon, A., Copeland, A. and McNally, P. (2018), "Empowering communities with data: role of data intermediaries for communities' data utilization", *Proceedings of Association for Information Science and Technology Annual Meeting*, Vol. 55 No. 1, pp. 583-592, available at, <https://doi.org/10.1002/pr2.2018.14505501063>

Further reading

- Association of Research Libraries (2017), "SPEC kit 354: data curation", available at: <http://publications.arl.org/Data-Curation-SPEC-Kit-354/> (accessed January 9, 2019).
- McKinsey and Company (2001), *Effective Capacity Building in Nonprofit Organizations*, Venture Philanthropy Partners, Reston, VA.
- McKinsey and Company (2003), "Why capacity building matters and why nonprofits ignore it", *New Directions for Philanthropic Fundraising*, No. 40, pp. 103-111, available at, <https://doi.org/10.1002/pf.36>
- Mosala, N.N. (2000), "Developing a capacity building policy for academic libraries", unpublished thesis, University of Cape Town, Cape Town, available at: <https://open.uct.ac.za/handle/11427/7816>

About the authors

Ayoung Yoon is Assistant Professor at Indiana University School of Informatics and Computing (IUPUI), Department of Library and Information Science. She was a RDA/US data share fellow for 2016–2017. She received a PhD Degree from the University of North Carolina at Chapel Hill and MSI from the University of Michigan. Her research areas include data curation, digital preservation and data sharing/reuse. She is particularly interested in building community capacity for constructing effective data infrastructure for sharing and reuse through proper curation practices. Currently, she is conducting several research projects focusing on communities' needs of data sharing, reuse and curation. Ayoung Yoon is the corresponding author and can be contacted at: ayyoon@iupui.edu

Devan Ray Donaldson is Assistant Professor in the Department of Information and Library Science, Indiana University, Bloomington (IUB). Donaldson is also Affiliated Faculty with the Data to Insight Center (D2I) at Indiana University. He is an internationally known Digital Curation Researcher. His research interests include digital repositories, data sharing practices, mass digitization, preservation management, preservation metadata, trust and security. His research has been funded by the University of Michigan, Indiana University, the Alfred P. Sloan Foundation, and the US Department of Energy. He received the PhD Degree in Information from the University of Michigan.

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com