The Realities of Research Data Management

PART FOUR

Sourcing and Scaling University RDM Services

Rebecca Bryant, Brian Lavoie and Constance Malpas

OCLC®

The Realities of Research Data Management

Part Four: Sourcing and Scaling University RDM Services

Rebecca Bryant
Senior Program Officer

Brian Lavoie
Research Scientist

Constance Malpas
Strategic Intelligence Manager & Research Scientist

OCLC®

April 2018

ORCID iDs
Rebecca Bryant iD https://orcid.org/0000-0002-2753-3881
Brian Lavoie iD https://orcid.org/0000-0002-7173-8753
Constance Malpas iD https://orcid.org/0000-0002-9312-8294

Please direct correspondence to:
OCLC Research
oclcresearch@oclc.org

Suggested citation:
Bryant, Rebecca, Brian Lavoie, and Constance Malpas. 2018. *Sourcing and Scaling University RDM Services*. The Realities of Research Data Management, Part 4. Dublin, OH: OCLC Research. https://doi:10.25333/C3QW7M.

CONTENTS

FIGURES

INTRODUCTION

In this report—the fourth and final of OCLC Research's *The Realities of Research Data Management* report series—we examine institutional choices for sourcing the provision, and scaling the deployment, of research data management (RDM) services.

By *sourcing*, we mean where RDM services are developed and managed: i.e., locally or by an external provider. By *scaling*, we mean at what scale will the services be deployed: i.e., at the level of the institution or at scales above or below the institution. In this report, we describe the sourcing and scaling choices made by our case study partners as they acquired RDM capacity and built their RDM service bundles.[1]

The Realities of RDM: A Brief Overview

The Realities of Research Data Management is a series of four reports looking at the context, influences, and choices research universities face in building or acquiring RDM capacity. Our findings are derived from detailed case studies of four research universities, hailing from four distinct national contexts: the University of Edinburgh (UK), the University of Illinois at Urbana-Champaign (US), Monash University (Australia), and Wageningen University & Research (WUR) (the Netherlands).

The first report in the series, *A Tour of the Research Data Management (RDM) Service Space*,[2] presents a simple framework for thinking about the RDM service space in its entirety (figure 1). The framework divides RDM services into three categories: *Education, Expertise, and Curation*. These categories summarize a wide array of specific services that may be deployed as part of a university's *RDM service bundle*—the range of RDM services offered by a university to its researchers. An RDM service bundle includes services that are built and deployed locally, as well as those that are sourced with external providers, with the university arranging access on behalf of its affiliated researchers.

RDM is not a monolithic set of services duplicated across universities. It is a customized solution shaped by a range of internal and external factors operating on local decision-making.

The second report, *Scoping the University RDM Service Bundle*,[3] examines in detail the choices our four case study partners made in selecting the set of services that would be included in their respective RDM service bundles. As we note in

Research Data Management Service Categories

EDUCATION
Raise awareness of RDM's importance, encourage RDM skill-building, and disclose RDM tools and resources

EXPERTISE
Decision support for, and customized solutions to, specific research data management problems

CURATION
Technical infrastructure and related services that support data management throughout the research cycle
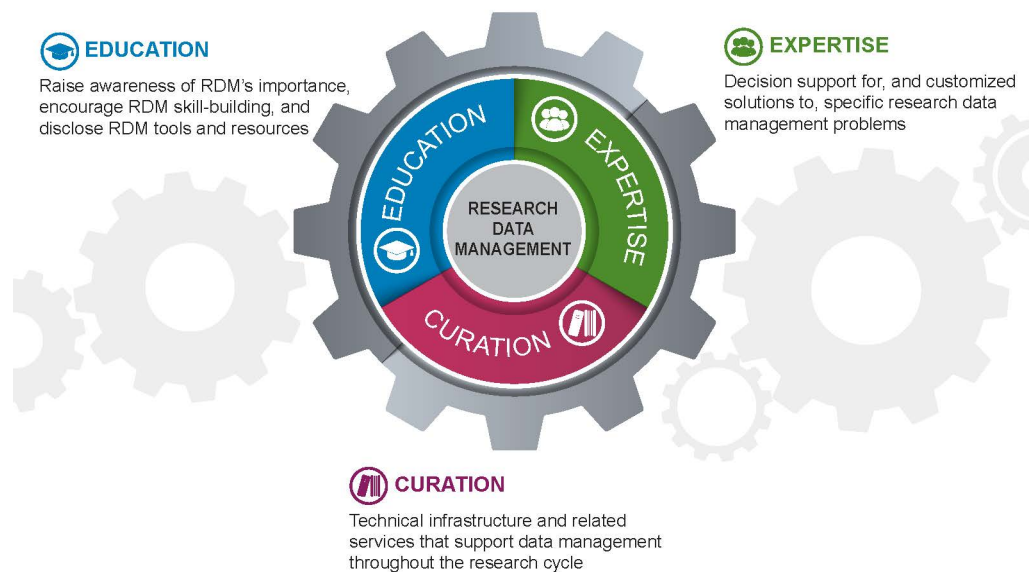
FIGURE 1. RDM SERVICE CATEGORIES FROM: *A TOUR OF THE RESEARCH DATA MANAGEMENT (RDM) SERVICE SPACE.* **THE REALITIES OF RESEARCH DATA MANAGEMENT, PART 1**[4]

www.manaraa.com

the report, a key takeaway from this analysis is that *RDM is not a monolithic set of services duplicated across universities. It is a customized solution shaped by a range of internal and external factors operating on local decision-making.* Each university selected RDM services in response to incentives or pressures emerging from both local circumstances and the broader environment in which it is situated. But what are these incentives that motivated universities to take action in the RDM space? This is the subject of our third report, *Incentives for Building University RDM Services*.[5] Based on our case studies, as well as the broader RDM landscape, we propose four broad categories of incentives potentially operating on a university's decision to develop an RDM service bundle: *compliance, evolving scholarly norms, institutional strategy, and researcher demand*. The key takeaway here is that rather than being some sort of scholarly fad, *RDM services are a response to real incentives that are driving university decision-making in this space. These incentives manifest differently in different university contexts, and they can change or evolve over time*. In this sense, RDM services will be sustainable and valued only to the degree they respond to these evolving incentives.

RDM services are a response to real incentives that are driving university decision-making in this space. These incentives manifest differently in different university contexts, and they can change or evolve over time.

Our previous reports focus on two key decision points in acquiring RDM capacity: deciding to act (responding to internal and external incentives to develop RDM services), and deciding what to do (scoping a bundle of RDM services for

deployment). In this, our final report, we examine a third decision point: once a university has decided to develop an RDM service bundle and has scoped the services it will contain, it must then determine *how* the capacity to support those services will be acquired and deployed. We explore the *sourcing and scaling* choices made by our case study partners, as well as frame the considerations that led to these choices. We conclude with some general insights about sourcing and scaling RDM services that emerge from our case studies.

Decision Point: *How* to Acquire RDM Service Capacity

A key finding of our earlier report, *Scoping the RDM Service Bundle*, is that no RDM service bundle is an island—all are connected, to a greater or lesser degree, to the broader, external RDM service ecosystem. More specifically:

> *… RDM services bundles are not self-contained. Although they differ in the degree to which they incorporate external services and resources, they are scoped to leverage some connection to the external RDM service ecosystem, creating a network of interdependence—varying in intensity from institution to institution—across the RDM service space.*

RDM service bundles emerge both from the ground—in the form of services developed, managed, and deployed locally—and from the cloud—in the form of services drawn down from the surrounding external ecosystem of RDM resources. An important task for university decision-makers is to choose a mix of internally and externally sourced RDM services appropriate for local needs and priorities. This is the choice we consider in this fourth and final installment of *The Realities of Research Data Management* series.

How to go about acquiring RDM capacity and establish services is a nuanced question, involving many considerations, such as:

- What local resources, in the form of staffing and infrastructure, are available?

- Are cooperatively sourced or nationally provisioned services an option?

- Is there a willingness to pay for commercially sourced solutions?

- Will capacity be built, bought, or licensed?

- Will the local RDM service bundle operate as a complement to, or in parallel with, other services in the broader RDM service ecosystem?

- Is the university interested in cultivating institutional prestige around locally built services?

Does the university source its RDM services locally or externally?

While in practice all of these questions and many more factor into the decision-making surrounding the implementation of RDM services, for our purposes we would like to reduce the complexities to a simple frame within which to examine the decisions made by our case study partners. Specifically, we express the issue as choices made along two dimensions: sourcing and scaling. Sourcing addresses the question: *does the university source its RDM services locally or externally?* In other words, are RDM services developed in-house, or is service provision outsourced? Framed in this way, sourcing choices address the question of *where* individual RDM services will be built, managed, and deployed, echoing the familiar business dichotomy "build or buy."

The terms "locally sourced" and "externally sourced" are general labels that obscure a great deal of nuance. An internally sourced RDM solution is not necessarily one that was built entirely from scratch—it could instead have been assembled from various components acquired elsewhere, such as open source applications, that are then integrated and adapted or customized to meet local requirements. Indeed, one of our case study partners—Edinburgh—adopted this approach in the development of some of its internally sourced RDM services. For the purposes of this study, we adopt an expansive view of internal sourcing that focuses on services that are built, customized, or adapted—and then deployed, maintained, and evolved—primarily through local effort and resources.

External sourcing of RDM services can take many forms, including a collaborative effort among peer institutions, commercial providers, non-profit organizations, national agencies, and even entrepreneurially minded universities. The RDM service ecosystem is becoming quite dense, and many universities will find they have many options in terms of potential external providers of various RDM services.

What is the scale of the user community that a particular RDM service is intended to serve?

In this report, we focus on sourcing questions regarding *university-supplied* RDM capacity. We recognize that RDM capacity can also be externalized by individual researchers constructing their own data management workflows from services available on the network, without the mediation of their university. This may include their use of external research data repositories for data sharing. We return to this topic briefly in the conclusion.

Scaling poses another choice for decision-makers: *what is the scale of the user community that a particular RDM service is intended to serve?* For example, are services intended for the general use of a wide spectrum of a university's affiliated researchers? Are they of a more specialized nature, addressing the requirements of specific research communities or disciplinary specialties within the campus? Or are they perhaps intended to serve a user community that extends beyond the boundaries of the campus, serving a cohort of scholars without regard to institutional affiliation?

Sourcing and scaling RDM services depend on a wide range of considerations, such as the prospect of cost savings from economies of scale, the desire to cultivate or enhance institutional reputation as an innovator, or to leverage or strengthen inter-institutional networks like consortia. Sourcing and scaling decisions are made at the level of an individual service, not the service bundle as a whole. As we have seen in our earlier reports,[6] the local RDM service bundle is often constructed as a mix of internal and external services, with the relative emphasis varying both in terms of the service category, and the sourcing and scaling philosophy of each university.

> *Sourcing and scaling decisions are made at the level of an individual service.*

The analysis that follows is organized according to the three service categories of the RDM service space illustrated in figure 1: Education, Expertise, and Curation. Each section discusses the experiences of our case study partners in one of these areas. The sourcing and scaling choices pertaining to Curation seem to offer the most fertile ground for analysis; indeed, most of the sourcing and scaling choices seem to be clustered here. We offer some observations on this in the conclusion of the report.

Sourcing and Scaling RDM Capacity: Case Studies

EDUCATION

Among the four institutions profiled in our case studies, educational resources for RDM are mostly locally sourced. Each of the four institutions offers a combination of self-directed orientation and learning resources (topical guides to RDM), instructor-led courses (in-person or virtual), and tools to assist students and faculty in developing basic RDM competencies. Each institution's Education service bundle for RDM includes components that are developed locally.

In all four case study institutions, locally developed resources for self-directed orientation and learning are provided by the university library. These guides supplement general information about university RDM services that are documented on the university website. We found no evidence that content for these guides had been sourced externally (e.g., licensed from a third party or cloned from another source). Library resource guides typically include references and links to external sources and the library resource guides on RDM followed the same general pattern, with many external links to further information and guidance.

> *The institutional effort that is devoted to creating and maintaining educational resources for RDM may be considerable.*

However, the guides offered by our case study partners are not simply lists of links, and typically include a general introduction to the principles and practice of RDM. In addition to general purpose "introduction to RDM" guides, some of the university libraries have integrated sections on RDM in disciplinary resource guides.[7] Thus the institutional effort that is devoted to creating and maintaining educational

resources for RDM may be considerable, even if it lends itself to some reuse—for example, including a general RDM section in the template for all disciplinary research guides.

Further evidence of direct, institutional investment in supporting RDM education may be found in the care that is put into organizing and aligning RDM guidance documents with researcher workflows. At the University of Edinburgh and Wageningen University, topical resource guides address data management needs at different stages of the research lifecycle, including project planning and grant proposal development, managing active data during a project, and preserving and sharing data after a project is complete.[8]

Beyond providing a conceptual framework within which to organize educational content, these design strategies help to position RDM services at the point of need and help to demonstrate that the university's RDM strategy is aligned with research workflows. This may help explain why universities invest local resources (time and effort) in producing and marketing educational resources for RDM, rather than outsourcing content creation; local customization assists in signaling the university's distinctive approach to supporting researchers.

The most resource-intensive approach to supporting RDM education is through in-person, instructor-led workshops. Here, too, each of our case study institutions is making significant investments in locally sourced RDM services. The scope of training varies, but some form of synchronous, in-person RDM training is offered by all four institutions. Wageningen University offers a one-day training course on research data management, organized by the Wageningen Graduate Schools in conjunction with the university library.[9]

At the University of Illinois, the Research Data Service offers five "Savvy Researcher" workshops focused on data management, on topics ranging from fundamentals of data management to understanding complex data

workflows.[10] Monash University Library offers a separate range of courses for research students and for university staff, and commits to customizing sessions according to the specific needs of different departments or research groups.[11]

And, while the University of Edinburgh can, in principle, outsource at least some of its educational programming in RDM to the UK Digital Curation Centre, it also offers a series of ten data management training courses staffed by university professionals.[12] These offerings supplement the self-paced MANTRA tutorial (billed as "a free online course for those who manage digital data as part of their research project) and a Coursera MOOC developed by the University of North Carolina-Chapel Hill in collaboration with the University of Edinburgh. While the MANTRA tutorial represents an Education service offering developed by EDINA and housed at the University of Edinburgh, it has become an important component of the shared infrastructure that not only benefits local Edinburgh users but also enables other institutions to externalize some RDM Education activity.

In some instances, as at Wageningen, education in the principles and practice of RDM has been integrated into the broader graduate educational curriculum. All doctoral researchers at Wageningen are required to submit a DMP as a part of their dissertation proposal; while enrollment in the RDM training workshop organized by the Graduate Schools and the university library is optional, there are clear incentives for student participation. Internalizing, or locally sourcing, RDM educational services is a strategic choice when it supports broader institutional policies (such as a university-wide DMP requirement) or educational objectives.[13]

If in-person, professional training in RDM is the costliest form of educational service a university can provide—as well as the most challenging to

outsource—training in the preparation of DMPs is arguably the easiest service to externalize. In North America and Europe, shared infrastructure has emerged in response to funder mandates for formal data management plans. The DMPTool provides customizable templates for institutions applying for grant funding from leading funders in the US; the DMPOnline provides customizable templates based on funder requirements in Europe.[14] A growing number of universities, including the University of Illinois and the University of Edinburgh, have established partnerships with DMPTool and DMPOnline, enabling them to customize templates for data management plans for institutional researchers.[15]

When outsourcing options for DMP preparation are not available, institutions may provide semi-automated support for creating data management plans. Wageningen University, which has a university-wide data management policy, provides a downloadable DMP template for individual researchers; completed forms are reviewed by graduate schools at the university before any research proposal is approved. Similarly, Edinburgh provides a generic DMP template for research projects that are not subject to a formal funder requirement. Monash University is the only institution among our case study partners that does not provide DMP templates as a service, although it does provide checklists of similar information.[16] There is neither a national funder mandate in Australia nor an institutional policy requirement for data management plans at Monash, which explains why the incentives for providing educational services for DMPs are so low.

EXPERTISE

Institutional RDM support is costly and providing Expertise services can be exceptionally challenging to institutions, as it requires a knowledge of data curation practices as well as expertise with software and domain-specific practices that can vary broadly across the heterogeneous research landscape.

In our study, we found that all four of our case study institutions provides local Expertise support, which could include individual consultation on tasks such as the preparation of DMPs, metadata creation, file storage and management, and occasionally even mediated deposit. Each institution provides researchers with a local email help line to request individualized support.

Expertise services require a "human layer" of knowledge.

Edinburgh and Wageningen rely primarily upon data curation staff members to work directly with researchers, making referrals as needed; Illinois has dedicated data curation staff as well as locally trained subject area librarians providing RDM expertise; and Monash relies exclusively on a distributed model. The Monash library serves as the central point of contact, and Monash researchers are referred to an array of campus units for RDM Expertise support, such as the eResearch Centre for advice on data storage and sharing, and the Records and Archives Service for advice on retention, appraisal, and de-accession.[17]

Expertise services require a "human layer" of knowledge—not just of metadata and information management practices, but also of heterogeneous disciplinary practices and software. This heterogeneity makes it impracticable for an institution to locally provide the needed expertise to curate the growing diversity of data. In response to this need, the University of Illinois, in collaboration with other North American partners and supported by an Alfred P. Sloan Foundation grant, is working to address "the challenge of scaling domain-specific data curation services and staff expertise collaboratively across a network of multiple institutions and digital repositories in order to provide expert data curation services in disciplines and domains beyond what any single institution might offer alone."[18]

The Data Curation Network (DCN) completed an initial pilot and published a report in July 2017 that outlines future plans for scaling Expertise capacity as well as learning, as DCN member institutions will not only share expertise but also engage together to develop skills and participate in a robust community of practice. It provides a framework for a cross-institutional staffing model that "seamlessly connects expert data curators to local datasets."[19]

> *A university is confronted with a fundamental choice: should curation systems be built and deployed locally, or should they be sourced from external providers?*

Libraries have a significant history of collaborating to solve problems of mutual interest in the digital space, evidenced by such initiatives as HathiTrust, DPN, and SHARE, and it seems likely that RDM will see more collaboration in time as the service space matures. The Data Curation Network is an example of institutions leveraging opportunities to move the provision of Expertise services to scales above the institution.[20]

CURATION

In our framework of RDM service categories, we define curation as the technical functions that ensure that research data sets are stored and managed in ways that promote ongoing integrity and accessibility. RDM Curation capacity involves a wide array of infrastructure and services necessary to meet these goals. In acquiring this capacity, a university is confronted with a fundamental choice: should curation systems be built and deployed locally, or should they be sourced from external providers?

Our case study partners represent a mix of solutions to this question. Figure 2 provides a lightweight visualization of the choices made by Edinburgh, Illinois, Monash, and Wageningen regarding sourcing RDM Curation infrastructure and services. In the figure, there are three sourcing options for Curation capacity: build it locally, source it within a cooperative effort of peer institutions, or externalize it to a third party provider, like a commercial service or national agency.

In figure 2, we represent the sourcing choices not as a set of discrete points, but as a continuum running from building locally at one end point to buying a service from an external provider at the other, with collaboration or partnership residing at the mid-point between the two. The idea is that there is no single form of building, partnering, or buying. In fact, many RDM capacity acquisition strategies tend to shade somewhere between these more starkly defined choices. We note some of the key features of the three basic models: building local systems maximizes the scope for control and customization over RDM services; partnering

Strategic Sourcing of RDM Capacity in Four Case Study Universities

Build
Maximize control, customization

Partner
Share development costs, scale expertise

Buy
Manage costs, relinquish some control

FIGURE 2. STRATEGIC SOURCING OF RDM CAPACITY IN FOUR CASE STUDY UNIVERSITIES

www.manaraa.com

allows for both cost sharing and pooling expertise among the collaborating institutions; and buying allows universities to manage costs by outsourcing service operation, maintenance, and development to an external party, at the expense of ceding some control over the current and future state of the service.

We have arranged our four case study partners in an order that parallels the sourcing continuum, suggesting a *very broad* characterization of their overall RDM Curation capacities. While these broad characterizations help to compare and contrast the strategies adopted by our partners for acquiring RDM Curation capacity, they also obscure a great deal of nuance and complexity in how these strategies are executed in practice.

As the figure indicates, Edinburgh is very oriented toward sourcing RDM Curation services locally. They see themselves as a pioneer in the RDM service space, in terms of building a comprehensive RDM service bundle, including customization of open source products like DSpace and OwnCloud. All of Edinburgh's major RDM Curation services, such as DataStore (a central file store supporting active data management), DataShare (an online repository of publicly discoverable data sets produced at Edinburgh), and DataVault (private archival storage for data sets with restricted access) were built and deployed locally.

An important reason for Edinburgh's adoption of local sourcing is their early entry into the RDM service space: as early adopters, they *by necessity* had to be early developers of RDM capacity, as few external options were available at the time. In this sense, Edinburgh's array of locally sourced RDM services are testimony to its longstanding commitment to research data management.

During our interview, RDM staff at Edinburgh noted that while the university has traditionally favored home-built solutions in areas like RDM, this ethos is beginning to shift, and there is a growing recognition that acquiring

capacity from outside sources may in some circumstances be a viable option. An interesting distinction was made between Curation services aimed at access, and those aimed at long-term preservation. For Edinburgh, preservation is a long-standing part of the institutional mission as it was expressed in the interview, Edinburgh intends to preserve research data in perpetuity, as it has been doing for other scholarly materials since 1583. On the other hand, if the service priority is current access and network visibility, then solutions outside the university, such as disciplinary repositories, may be appropriate.

Illinois shows a similar, yet slightly different pattern, as Edinburgh. Nearly all of Illinois's RDM Curation services are locally sourced—in particular, the Illinois Data Bank, a public access repository for research data from Illinois researchers.[21] One rationale for adopting a locally sourced approach was that doing so allowed the university to achieve progress in development and deployment on its own timeframe.

A complex, lengthy university procurement process—extending to as much as a year and a half—was also mentioned as a factor, as well as concern over some of the prevailing pricing models for external solutions. For example, some pricing models are based on the total FTE at an institution, even though only a portion of the university community was likely to utilize the service. Finally, there was some concern that external providers may not be adequately responsive to university needs, especially given Illinois's local digital preservation infrastructure is growing in maturity and strength.

Despite these considerations, Illinois RDM staff view the strategy of relying on locally sourced RDM solutions more as a current necessity than a desired outcome. Ideally, Illinois would like to move toward more cooperatively sourced services deployed at group scale, as it has already begun to facilitate Expertise sharing through the Data Curation Network. While Illinois would like to see more RDM

services move above the institution—perhaps to the consortial level—they recognize that there are obstacles to overcome to achieve this. For example, universities will tend to differ in the specific RDM requirements they would expect a collaboratively sourced system to support, making agreement on the system's specifications difficult to achieve. In the meantime, because of the resources allocated by the Illinois Office of the Vice Chancellor of Research, it was important that progress be demonstrated in a timely manner.

In contrast to Edinburgh and Illinois, Wageningen has adopted a group-scale approach to sourcing RDM Curation services. Rather than building RDM Curation services locally, they rely on an external eco-system of RDM services deployed at both the consortial and national scale. A key rationale for this approach was that Wageningen takes the view that building and managing data curation infrastructure requires special expertise that the university has not yet developed internally; given this, a better solution is to rely on externally provided Curation services. These services include 4TU.ResearchData, a repository serving the four members of the 4TU consortium (an alliance of Dutch technical universities, of which Wageningen is the most recent member), as well as the national-scale DANS-EASY data repository, provided by the Dutch Data Archiving and Network Services (DANS) institute, which offers consultation and services aimed at supporting permanent access to digital resources.

It is interesting to contrast Wageningen's experience to that of Edinburgh. As noted, Edinburgh attributed part of its decision to source locally to the fact that it was an early entrant to the RDM service space, when few external options were available. In contrast, Wageningen acquired RDM Curation capacity relatively recently and was able to utilize two existing repository services. Wageningen staff also noted the advantages of working within a relatively small Dutch higher education system

in terms of above the institution cooperation: "everyone knows everyone else," and, as a result, trust networks are strong, which in turn facilitates inter-institutional collaboration around services such as the 4TU data repository. In the same way, the geography over which Dutch higher education institutions are distributed is relatively compressed—universities are not too far distant from one another—which also helps to ease some of the obstacles of deep collaboration around RDM.

Monash represents still another approach to acquiring RDM capacity. Monash's RDM Curation services are distributed across a range of external providers, including commercial, regional, and national services. The centerpiece of Monash's RDM Curation capacity is monash.figshare, an instance of the commercially provided figshare for institutions, a customizable portal for institutional research outputs; this service acts as an interface and uploading process to data storage resources coordinated by Monash. The storage resources include VicNode, serving the state of Victoria as part of the storage node network operated by the Australian government's Research Data Services. In addition, Monash uploads metadata from monash.figshare to Research Data Australia, a research data discovery service offered through the Australian National Data Service (ANDS).

Monash's decision to adapt an externally based eco-system of RDM Curation solutions stems from an internal "buy first" philosophy: services should be developed internally only when absolutely necessary. From the institution's perspective, internal development and technical staff were scarce, with little excess capacity to support new RDM systems. Open source solutions were not a good option for the same reason, and were regarded as expensive to maintain. Another consideration was a lack of conviction that it was a good strategy for Monash to be in the "data storage business"—it could never compete with the level of investment and resources of companies like Amazon.

This points out a distinctive feature of Monash's RDM capacity: its inclusion of commercially provided services. Monash staff did note that ANDS is where they usually look for centralized services, although sometimes Monash's needs and requirements are too different or specific. They report being pleased with their relationship with figshare, and, in particular, have appreciated its responsiveness—although they acknowledge the risk that as their commercial partner grows, it may need to support some requirements that are not a priority for Monash and responsiveness may diminish.

While it is easy to broadly categorize the four case study partners in regard to their sourcing choices—Edinburgh and Illinois adopting a locally sourced approach, with Wageningen and Monash pursuing an externally focused strategy—a closer look at the sourcing for the individual services comprising each university's RDM Curation capacity suggests that none of our case study partners fit precisely in a local or external model. Rather, each adopts a hybrid model mixing local and external sourcing choices, with a discernable emphasis on local or external sourcing emerging at the level of the overall RDM Curation service bundle.

A distinctive feature of Monash's RDM capacity is its inclusion of commercially provided services.

While exhibiting an overall focus on internal or external sourcing of RDM Curation services, each of our case study partners stitches together an RDM Curation capacity assembled from both internal and external sources. For example, Edinburgh is perhaps the "purest" example among our case studies of a strategy for acquiring RDM Curation capacity that relies on local adaptation of open source solutions. Indeed, most of the major Curation services offered by Edinburgh follow this model. And yet, we can still find examples in the Edinburgh Curation service bundle of

services sourced elsewhere: for example, their instance of the Pure research information management (RIM) system is used to construct a registry of data sets produced by Edinburgh researchers, whether archived locally or outside the university; this system is sourced with a commercial provider, Elsevier. It also brokered a bulk licensing of the RSpace Electronic Lab Notebook (ELN) platform to support Edinburgh researchers.[22]

Edinburgh is perhaps the "purest" example among our case studies of a strategy for acquiring RDM Curation capacity that relies on local adaptation of open source solutions.

Similarly, while Illinois opted to source its data repository locally, it supports private or restricted sharing of data among research collaborators through U of I Box, an Illinois-branded instance of the cloud-based content management/file sharing service Box. Wageningen, while looking outside the university for data repository services, nevertheless manages Git@WUR, a local instance of the GitLab source code management application. And Monash, which like Wageningen utilizes an externally focused strategy for RDM Curation services, has developed the MyTardis data management platform, which supports the capture and storage of data from scientific instrumentation; Monash manages a local implementation of this system called Store.Monash.[23]

Internal or external sourcing is the fundamental choice when acquiring RDM Curation capacity. However, the scale of the RDM offering—whether internally or externally sourced—is also an important decision point. When we think about RDM Curation services, it is common to view the institution as the typical scale of deployment: the university acquires RDM capacity on behalf of its affiliated researchers.

www.manaraa.com

Indeed, this is often the scale that universities deploy RDM services in practice. However, we cannot overlook that in some cases, differences in data management practices and requirements across disciplinary or research communities can result in RDM solutions emerging at scales *below the institution*: that is, for specific groups within the campus. And this can be the catalyst for larger, institution-scale deployments. Similarly, a university's RDM solution may be intended for user communities scaled *above the institution*: that is, communities that extend beyond the university's affiliated researchers.

We see examples of each of these scenarios in the context of our case study partners. Some of Monash's earliest efforts in RDM began with solutions customized for local researchers working in protein crystallography. As Monash staff noted, this turned out to be important engagement, because it demonstrated both a concrete RDM need that the university could fulfill, and was useful proof of concept work for more general RDM services that would follow.

Similarly, Wageningen's efforts to acquire RDM capacity were catalyzed by the Graduate Schools, which sought training for graduate students in data management and the development of DMPs, with further services (including curation) to be developed based on the needs and requirements arising out of what was specified in DMPs. In the case of both Monash and Wageningen, services developed initially for circumscribed groups on campus eventually evolved into investment in campus-wide RDM solutions.

While Monash and Wageningen underscore the importance of "below the institution" approaches to scaling RDM services, Edinburgh is an example of an institution scaling RDM services beyond the boundaries of the university. As mentioned earlier, Edinburgh sees itself as a leader in the RDM service space; in light of this, it perceives a role for itself in supporting the wider scholarly community in regard to RDM

services. For example, Edinburgh was a participant in the consortium that launched the UK's Digital Curation Centre (DCC) in 2004, and now serves as the DCC's host institution.

In addition, Edinburgh has submitted a bid to become an RDM service provider through Jisc's Research Data Shared Service project, an initiative currently under development to create a "framework of suppliers" of RDM-related services to UK higher education institutions. In short, Edinburgh's RDM strategy is leading it to be seen not just as a provider of RDM services for its affiliated researchers, but as a global center of excellence in the RDM service space.

Conclusions

Our four case studies have provided useful insights on how universities in different national contexts have addressed questions about what services to provide, how to source their services, and resource requirements. Each of our case study partners has pursued a different strategy, summarized as:

- Edinburgh favors home-built, locally sourced solutions, often based on open source foundation

- Illinois has also adopted this strategy, but is seeking opportunities for "above the institution" solutions

- Wageningen has outsourced much of its RDM Curation capacity to existing consortial and national scale solutions, while internalizing Education and especially Expertise services

- Monash has adopted an externally focused approach incorporating regional and national-scale solutions, as well as a commercial provider, and provides local Education and Expertise through distributed local services

CURATION SERVICES ARE THE MOST LIKELY TO BE EXTERNALIZED

Looking at these case studies, and considering the broader RDM ecosystem, we observe that Curation activities appear to be the most amenable to externalization. Why? First and foremost, because the technical infrastructure required for curation is expensive to develop or manage locally, the incentives to leverage shared infrastructure are strong. Furthermore, the infrastructure that is needed for managing and preserving research data is well suited to web-scale solutions, as evidenced by the proliferation of external solutions in the current environment (e.g., national agencies like DANS, disciplinary data repositories like Dryad, and commercial services like figshare).

> *Curation activities appear to be the most amenable to externalization.*

And, because the market for shared curation is relatively strong, commercial and consortial providers are motivated to develop solutions. Finally, while institutional norms (e.g., a conviction that locally developed, bespoke solutions are preferred) may produce some drag on externalization, those norms are less likely to influence decisions about how to source Curation capacity than decisions about Education or Expertise. This is because Education and Expertise assist in deepening university engagement with researchers, and may actually contribute to institutional reputation.[24]

In short, the incentives to externalize Curation are stronger than the incentives to external Education or Expertise.

In the cases we examined, the propensity to externalize RDM Curation services rests in part on the availability of "above the institution" RDM Curation services from external providers like national agencies and commercial entities. The availability of shared infrastructure for research data management (and other university needs) varies considerably in different regions and higher education systems.

In places like Australia, the Netherlands, and the UK, where most funding for higher education and research is centralized through the national government, a variety of shared services have emerged to support university-based teaching and learning. Institutions in the Netherlands can source RDM services through the national DANS EASY[25] data repository; in Australia, universities may rely on the Australian National Data Service (ANDS)[26] for national-scale research data discovery services; in the UK, there is ongoing discussion of a shared services model for RDM. In these settings, the decision to externalize, or internalize, components of the RDM service bundle is a matter of strategic choice.

This may be of particular relevance for relative newcomers to the RDM service space, who may have the option to utilize a range of established, externally sourced RDM services that were not available to universities that were among the earliest to acquire RDM capacity. It is also significant to institutions in the US and other places (including in the UK where shared services for RDM are under consideration, but not yet available services) where there is relatively little national-scale infrastructure for university-based research. As noted, the decision to source RDM Curation services internally by both Edinburgh and Illinois was at least partially determined by a lack of viable external options. Yet, as the existence of the shared 4TU Data-Centre[27] in the Netherlands shows, the existence of national-scale infrastructure does not necessarily constrain the development of complementary consortial infrastructure in the same service space.

In examining the Curation capacity developed by universities, it's important to acknowledge that institutional Curation activities exist alongside an extensive array of disciplinary data repositories

www.manaraa.com

and other RDM resources and services that researchers procure for themselves, without university intervention. Put another way, the existing RDM workflows established by individual researchers may be instrumental in convincing a university to eschew building local capacity. Several of our case study partners mentioned that they recognize that some research groups on campus were fulfilling their RDM requirements via external, disciplinary-focused RDM services; the universities accept this reality and are content to let it continue, rather than attempt to build competing local capacity and/or entice researchers away from established data workflows that involve services beyond the university's control.

As Monash University's figshare implementation demonstrates, adoption of commercial RDM solutions can co-exist alongside implementation of shared public infrastructure. Indeed, some commercial solutions rely on services provided by national agencies: DANS in the Netherlands provides back-end data archiving services to the commercial Mendeley Data platform, which supports researcher deposit of data sets and is licensed by some universities as a component of the local RDM service bundle.[28]

> *Existing RDM workflows established by individual researchers may be instrumental in convincing a university to eschew building local capacity.*

"Buying into" commercial solutions for one component of RDM service provision does not preclude, and may indeed facilitate, the development of innovative local solutions for some other component. Consider the example of Wageningen, where selective outsourcing of Curation capacity has enabled the university to differentiate itself as a center of excellence in RDM expertise.

Of course, a university's strategic sourcing decisions—for research data management, campus email systems, or even catering services—are shaped by perceptions about the optimal scale of certain operations, as well as the availability of external solutions. Despite the potential cost efficiencies of externalizing university data management to commercial cloud storage providers, many universities prefer to manage enterprise data archives locally, for example. Research data is an institutional asset, which may have tangible business value (in the form of patents, for example) or long-term strategic value (developing a center of excellence in fisheries science, or humanities computing services).

Additionally, general concerns about the benefits and tradeoffs of outsourcing management or preservation of locally created research outputs (or even metadata describing those outputs) may outweigh cost efficiencies of partnering with commercial partners offering 'cloud' data hosting services. Some universities will choose to externalize research data discovery and archiving services to national or commercial providers, while others will prefer hybrid solutions with "best of breed" components sourced selectively from a mix of internal and external sources.

EDUCATION AND EXPERTISE SERVICES ARE LARGELY LOCAL IN SCALE . . . SO FAR

Conversely, there are incentives for institutions to develop Education and Expertise services locally. As we articulated in our third report, local educational outreach may be necessary to increase research awareness and engagement, encouraging researchers to use newly offered Curation services. It is also true that institutions need to allocate local human resources with the knowledge to lead RDM initiatives, make programmatic decisions, develop resources, and train others to provide services.

Because we have defined Expertise services as being primarily human-mediated, this almost

necessarily favors local sourcing over external sourcing, because the one-on-one interaction between expert/consultant and researcher is facilitated by co-location on campus. This local Expertise is important for establishing local credibility with researchers, contributes to institutional differentiation, and likely accounts for the rapid emergence of "data librarian" roles in our case study institutions.

There are incentives for institutions to develop Education and Expertise services locally.

As we've seen through this tour of the RDM service space, there is no shortage of external, group-scale resources to support Education, which research universities may leverage in compiling their RDM service bundle, such as:

- The highly-rated Coursera course developed by librarians at the University of North Carolina at Chapel Hill and the University of Edinburgh.[29]

- Adaptation of LibGuides and other web-based resource guides, sometimes cloned (with or without acknowledgement) from other LibGuide sources.[30]

- Some publicly funded and/or subsidized organizations, like the UK Digital Curation Centre (DCC) and the European Association of Databases for Education and Training (EUDAT) Collaborative Data Infrastructure, offer fee-based educational offerings on a cost-recovery basis.[31]

- Data Carpentry, a non-profit organization, provides fee-based workshops on a range of data literacy topics in select disciplines,

with an emphasis on increasing researcher competency in active data management.[32]

Notwithstanding the availability of these exemplars of group-scale Education services, and the fact that almost any university can source or supplement local training in RDM through low- or no-cost online offerings, our case study institutions are not primarily reliant upon these. Instead, they choose to dedicate local resources to interact with, educate, and guide local researchers through the development of local policies, workshops, courses, and online resources.

Today institutions are also primarily delivering Expertise services at the local scale. We believe that the provision of a minimal level of local services, particularly Expertise, is necessary for institutions to establish credibility—both with their local researchers and also within the larger RDM scholarly communications community.

However, we expect to see growth in group-scale support for RDM Expertise services, driven primarily by the heterogeneity of disciplinary practices that make it impracticable for even the largest, most well-resourced institutions to locally provide all the needed domain and technical expertise, such as the following:

- DataQ, a collaborative platform developed by a partnership of US libraries, functions as a virtual call center for questions about research data, scaling RDM expertise to the academic library community at large. Individual researchers (or librarians) can post questions to a message board; responses are crowd-sourced from a network of expert volunteers with oversight from an editorial board.[33]

- Formalized RDM expertise-sharing networks such as the US Data

Curation Network (DCN), Portage network of experts (Canada), and the Netherlands' National Coordination Point for Research Data Management (LCRDM) are further examples of efforts to leverage distributed expertise at a group-scale capacity.[34]

As a result, we expect a future in which local RDM Education and Expertise services are offered in combination with "above the institution" solutions. Given the breadth of the market as well as the growing uniformity of some curricular content for RDM practice, there will be future opportunities for RDM to scale efficiently, especially in the digital environment, and as they are widely shared and increasingly standardized by funder mandates.

> *We expect a future in which local RDM Education and Expertise services are offered in combination with "above the institution" solutions.*

In considering sourcing choices for RDM services—whether Education, Expertise, or Curation—a key factor that must be accounted for is uncertainty. Nearly all of our case study partners cited uncertainty about the future of the RDM service environment as a key factor impacting their sourcing decisions, and this is likely to hold for many other universities as well. The RDM service space is still quite fluid, and both the nature of the services needed to support data management, and where those services are best sourced, are still very much unsettled. For example, while universities may be interested in externalizing certain RDM services to a collaborative effort of peer institutions, those services may be currently operated through grant funding of finite duration. Without a sufficiently robust business model to sustain them, the long-term future of services

dependent upon finite grants is uncertain. This introduces a significant risk into the externalization choice: should universities rely on services and infrastructure that may not persist? Uncertainty also impacts the choice to develop internal solutions. It may not be prudent to invest heavily in locally sourced RDM services and infrastructure if there is a strong probability that a reliable external sourcing option may soon be available at greater scale and lower cost. Making strategic choices regarding internal vs. external sourcing of RDM services remains a fraught exercise in a service space that is still relatively immature and dynamic.

ONE SIZE DOES *NOT* FIT ALL: SOURCING AND SCALING RDM TO FIT INSTITUTIONAL NEEDS

For research-intensive institutions like our four case study partners, providing a robust and comprehensive RDM service bundle confers both reputational and operational benefits— enabling data-intensive research activity in a variety of disciplines. That benefit is maximized when some part of the RDM service offering is internalized as a distinctive, value-generating university activity. Selective externalization of other components may be part of a broader strategy to redirect institutional resources (financial investments, staff attention, etc.) toward activities that deliver a greater benefit to the university.

> *It would be a mistake to imagine that there is a single, best model of RDM service capacity, or a simple roadmap to acquiring it.*

For example, at Wageningen, a strategy of outsourcing some Curation activities has enabled deeper specialization in Expertise services that help to distinguish the university as a center of excellence in data-driven agronomy and animal sciences. In contrast, at Edinburgh,

a strategy of internally sourcing Education, Expertise, and Curation components has established the university as a global leader in RDM and a potential service provider to other institutions in the UK.

It would be a mistake to imagine that there is a single, best model of RDM service capacity, or a simple roadmap to acquiring it. However, most universities that support a large-scale research enterprise will likely need to develop or acquire institutional RDM capacity in Education, Expertise, and Curation, but only those with entrepreneurial ambitions or reputational stake in RDM-related innovation are likely to pursue a strategy of deliberate internalization for the long term. Less research-intensive institutions may also require some local RDM services, but they are likely to look somewhat different. These colleges and universities may find that the appropriate value-maximizing strategy for

acquiring RDM capacity will rely on group-sourced and commercial solutions that do not require substantial local customization.[35]

In our view, a "minimalist" approach to local RDM service provision is no worse (or better) than a "maximalist" approach, provided it aligns with institutional needs. Decisions about whether to build or buy RDM capacity, and choices about the optimal scale of deployment, will be informed by larger institutional interests. Is the university actively seeking to increase its research reputation in data-intensive fields? Or pivoting from an emphasis on liberal education to career-directed learning? Does the university aspire to be a center of technical innovation in data management? The benefits and tradeoffs of internalizing or externalizing RDM capacity, and of scaling its deployment below, at, or above the institution will differ in each case.

ACKNOWLEDGMENTS

NOTES

1. We are pleased to acknowledge our colleague Lorcan Dempsey, who introduced the term "Sourcing and Scaling" in a 2010 blog post. See *Lorcan Dempsey's Weblog*, posted 21 February. http://orweblog.oclc.org/sourcing-and-scaling/. This blog post is also included in Lorcan's later book *The Network Reshapes the Library:* See Dempsey, Lorcan. 2014. *The Network Reshapes The Library: Lorcan Dempsey on Libraries, Services, and Networks,* edited by Kenneth J. Varnum. Chicago: ALA, American Library Association, 41-43. https://www.oclc.org/content/dam/research/publications/books/dempsey-network-reshapes-library.pdf.

2. Bryant, Rebecca, Brian Lavoie, and Constance Malpas. 2017. *A Tour of the Research Data Management (RDM) Service Space*. The Realities of Research Data Management, Part 1. Dublin, Ohio: OCLC Research. https://doi.org/10.25333/C3PG8J.

3. Ibid.

4. Bryant, Rebecca, Brian Lavoie, and Constance Malpas. 2017. *Scoping the University RDM Service Bundle.* The Realities of Research Data Management, Part 2. Dublin, OH: OCLC Research. https://doi.org/10.25333/C3Z039.

5. Bryant, Rebecca, Brian Lavoie, and Constance Malpas. 2018. *Incentives for Building University RDM Services.* The Realities of Research Data Management, Part 3. Dublin, OH: OCLC Research. https://doi.org/10.25333/C3S62F.

6. See note 4.

7. The University of Illinois offers an "Introduction to Data Management for Undergraduate Students" resource guide (http://guides.library.illinois.edu/introdata) and also incorporates information and guidance on RDM in specialist guides, such as "Neurosciences Research and Scholarship" (http://guides.library.illinois.edu/c.php?g=425289&p=3168341). Similarly, Monash University offers a general orientation to RDM in the "Graduate Research: Managing Research Data" guide (http://guides.lib.monash.edu/gradres/research-data) and integrates information about campus RDM services within disciplinary guides like this guide for "Maintenance and Reliability Engineering" http://guides.lib.monash.edu/c.php?g=219722&p=1452925.

8. At Wageningen, RDM guidance is grouped into three research lifecycle categories: planning, doing, and finishing research projects, and a fourth category related to compliance with university data policies. See "Data Management." https://www.wur.nl/en/Expertise-Services/WDCC/Data-Management-WDCC.htm.

9. See "Research Data Management." Course Schedule. https://wgs.crs.wur.nl/courses/details/84/.

10. See "Research Data Service." Workshop. https://www.library.illinois.edu/rds/workshops/.

11. See "Research Data Skills Development." Sessions for Research Students Currently offered. https://www.monash.edu/library/researchdata/skills.

12. See "Information Services: Research Data Service." Data Training. https://www.ed.ac.uk/information-services/research-support/research-data-service/training.

13. Jacquelijn Ringersma from WUR provides an in-depth description of this policy and its creation in an OCLC Research webinar, "Policy Realities in Research Data Management," posted 25 July 2017. https://www.oclc.org/research/events/2017/07-25.html.

14. Both tools provide free access to templates for individuals at non-partner institutions. DMPTool is a service of the University of California Curation Center of the California Digital Library. See: https://dmptool.org/about_us. The DMPonline service is supported by the UK Digital Curation Centre. See: https://dmponline.dcc.ac.uk/. DMPonline has been customized for use by Canadian research organizations by the Portage data stewardship network, supported by the Canadian Association of Research Libraries/Association des bibliothèques de recherche du Canada. See: https://assistant.portagenetwork.ca/.

15. DMPTool and DMPOnline each report having approximately 200 institutional partners. See: https://dmptool-stg.cdlib.org/public_orgs, and https://dmponline.dcc.ac.uk/. [requires sign in]

16. Monash University. "Data Planning." In Library: Managing Research Data. https://www.monash.edu/library/researchdata/guidelines/data-planning.

17. Monash University. "Advice and Planning." In Library: Managing Research Data. https://www.monash.edu/library/researchdata/advice.

18. Johnston, Lisa R, Jake Carlson, Cynthia Hudson-Vitale, Imker, Heidi; Kozlowski, Wendy; Olendorf, Robert; Stewart, Claire. 2017. *Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data.* Minneapolis, MN: University of Minnesota's Digital Conservancy, *8*. http://hdl.handle.net/11299/188654.

19. Ibid., 4.

20. Dempsey, Lorcan, Constance Malpas, and Brian Lavoie. 2014. "Collection Directions : Some Reflections on the Future of Library Collections and Collecting." Libraries and the Academy 14,3 (July): 393–423. http://muse.jhu.edu/journals/portal_libraries_and_the_academy/summary/v014/14.3.dempsey.html, Portal Preprint: http://www.press.jhu.edu/journals/portal_libraries_and_the_academy/portal_pre_print/articles/14.3_dempsey.pdf, OCLC Research Preprint [see page 30]: http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-collectiondirections-preprint-2014.pdf; Dempsey, Lorcan. 2018. "The Powers of Library Consortia 2: Soft Power and Purposeful Mobilization: Scaling Learning and Innovation." *Lorcan Dempsey's Weblog*, Posted 6 March 2018. http://orweblog.oclc.org/the-powers-of-library-consortia-2-soft-power-and-purposeful-mobilization-scaling-learning-and-innovation/.

21. Fallaw, Colleen, Elise Dunham, Elizabeth Wickes, Dena Strong, Ayla Stein, Qian Zhang, Kyle Rimkus, Bill Ingram, and Heidi J. Imker. 2018. "Overly Honest Data Repository Development." *Code4Lib Journal* 39 (February). http://journal.code4lib.org/articles/11980.

22. In fact, three of our four case study institutions (Edinburgh, WUR, and Monash) utilize Elsevier's Pure research information management (RIM) system to collect the institutional record of research data sets locally created, even if not locally curated. For information about the RSpace licensing at Edinburgh, see https://www.ed.ac.uk/information-services/research-support/research-data-service/working-with-data/eln.

www.manaraa.com

23. At the time of our interview, Wageningen was also investigating active data management requirements at the university. Staff noted that they will likely look at internal capacity for active data management, and continue to rely on external solutions for archiving data once the research process has been completed.

24. Elsewhere, our colleague Lorcan Dempsey has explored the importance of transaction costs in defining the boundaries of library service and the incentives to externalize various services. See, for example, Dempsey, Lorcan. 2012. *Libraries and the Informational Future: Some Notes* (revised 08 August), especially chapters 10-14. https://www.oclc.org/content/dam/campaign-landing-pages/en/linked-data/dempsey-informationalfutures.pdf, [originally published in *Information Professionals 2050: Educational Possibilities and Pathways.* Edited by Gary Marchionini and Barbara B. Moran. Chapel Hill, NC: University of North Carolina at Chapel Hill.113-127. https://sils.unc.edu/sites/default/files/news/Information-Professionals-2050.pdf.]

25. See https://easy.dans.knaw.nl/ui/home.

26. ANDS "Developing a National Research Data Cloud." https://www.ands.org.au/.

27. See http://researchdata.4tu.nl/home/. [in Dutch]

28. See DANS (Data Archiving and Networking Services). 2015. "Collaboration DANS and Mendeley on Archiving Datasets Via the New Data Repository Platform Mendeley Data." *News and Events*, 23 September 2015. https://dans.knaw.nl/en/current/news/collaboration-dans-and-mendeley-on-archiving-datasets-via-the-new-data-repository-platform-mendeley-data.

29. See https://www.coursera.org/learn/data-management. This five-week course provides an introduction to core concepts in research data management; for a modest fee, online learners can earn a Certificate of Completion attesting to their competency in the subject.

30. See, for example, the Data Management guide at the University of Central Florida, which acknowledges external sources: http://guides.ucf.edu/c.php?g=78715&p=504529; Duke University Library's Research Data Management resource guide includes several sections designed for re-use by liaison librarians in subject-specific LibGuides: https://guides.library.duke.edu/c.php?g=633433&p=4771399.

31. The Data Curation Center provides training, and consultancy is offered on a cost-recovery basis: http://www.dcc.ac.uk/training/request-training-session. The European Union's Horizon 2020 program, the European Association of Databases for Education and Training (EUDAT) Collaborative Data Infrastructure, provides training on RDM to research organizations throughout Europe, primarily through online course modules: https://www.eudat.eu/training

32. Data Carpentry currently offers workshops and course materials in ecology and genomics. Workshops for other disciplines are under consideration and/or in development. Course materials are distributed under a Creative Commons license to maximize reuse. The organization is supported by grant funding from the Gordon & Betty Moore Foundation, and membership dues from affiliated organizations. Universities often host the workshops and make them available to students and faculty at no cost; to date, such workshops have been organized in 23 countries worldwide. See: http://www.datacarpentry.org/about/.

33. See http://researchdataq.org/index.php.

34. The Data Curation Network "aims to "develop a 'network of expertise' model for data curation services that will enable academic libraries to collectively, and more effectively, curate a wider variety of data types—for example, by discipline and file format—that expands beyond what any single institution might offer alone." See: https://sites.google.com/site/datacurationnetwork/; The Portage Network of Expertise is comprised of eight expert groups focused on Data Management Planning, Preservation, Discovery, RDM Training, Research Intelligence, Curation, Research Data and Ethics, and Peer Certification of Data Repositories. See: https://portagenetwork.ca/about/network-of-expertise/. The Netherlands National Coordination Point for Research Data Management includes five working groups focused on Awareness and Engagement, Facilities and Data Infrastructure, Financial Aspects, Legal Aspects, and Research Support and Advice. See: https://www.edugroepen.nl/sites/RDM_platform/SitePages/Home.aspx. [in Dutch]

35. As noted above, the University of Illinois's decision to internalize some curation capacity was driven in part by a concern that FTE-based pricing of commercial services would *over-estimate* the value of service to the university as a whole. Service-providers in the RDM space (and elsewhere) might find that market opportunities would be greater, if pricing and business models acknowledged that the value an institution can derive from a product will be impacted by the mix of activities that are important to the university. Even research-intensive universities are balancing investments in teaching, learning, and service against investments in the research enterprise.

For more information about OCLC Research's work on research data management, please visit: **oc.lc/rdm**

OCLC®