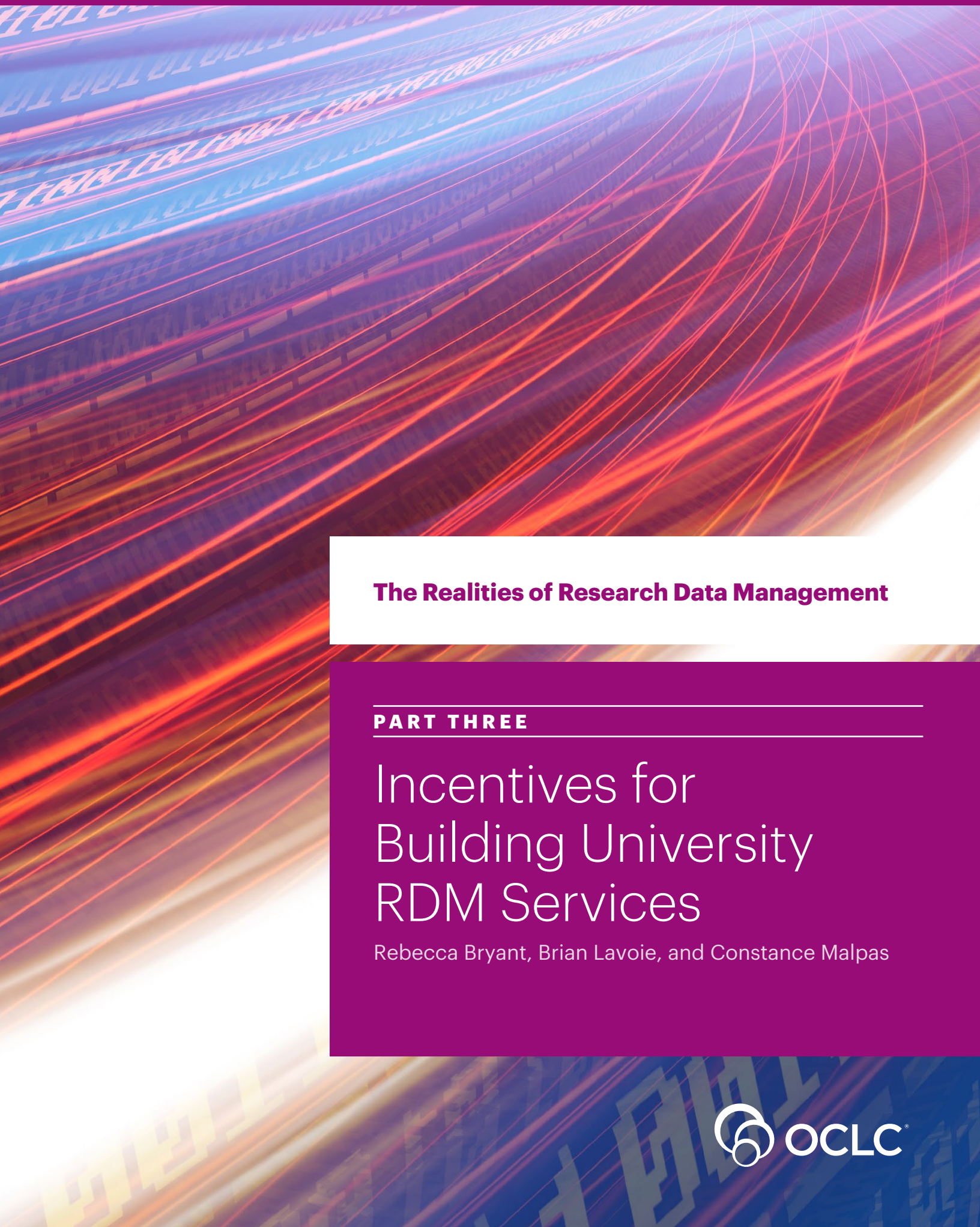
OCLC RESEARCH REPORT



The Realities of Research Data Management

PART THREE

Incentives for Building University RDM Services

Rebecca Bryant, Brian Lavoie, and Constance Malpas

OCLC

The Realities of Research Data Management

Part Three: Incentives for Building University RDM Services

Rebecca Bryant
Senior Program Officer

Brian Lavoie
Research Scientist

Constance Malpas
Strategic Intelligence Manager & Research Scientist

January 2018

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

ISBN: 978-1-55653-043-2
DOI: 10.25333/C3S62F
OCLC Control Number: 1017698241

ORCID iDs
Rebecca Bryant https://orcid.org/0000-0002-2753-3881
Brian Lavoie https://orcid.org/0000-0002-7173-8753
Constance Malpas 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. *Incentives for Building University RDM Services*. The Realities of Research Data Management, Part 3. Dublin, OH: OCLC Research. doi:10.25333/C3S62F.

CONTENTS

FIGURES

INTRODUCTION

The Realities of Research Data Management is a series of reports by OCLC Research that examines the context, incentives, and choices made by research universities related to research data management practices, support, and capacity. Our findings are derived from detailed case studies of four research universities, hailing from four distinct national contexts: the University of Edinburgh (Scotland); the University of Illinois at Urbana-Champaign (USA); Monash University (Australia); and Wageningen University and Research (Netherlands).

Our first report,[1] *A Tour of the Research Data Management (RDM) Service Space*, provides background context and a framework for subsequent reports. The second report,[2] *Scoping the University RDM Service Bundle*, closely examined how each institution scoped its local RDM services, concluding that RDM services are not a monolithic set of services duplicated across universities, but are instead customized in response to local and external circumstances.

After richly documenting the "what" in RDM service provision at our four institutions in the first two reports, this third report in the series provides a complementary in-depth exploration of the internal and external incentives that influenced decision making, priorities, and services offered.

The RDM Service Space

The first report in this series, *A Tour of the Research Data Management (RDM) Service Space*,[3] 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 local RDM services offered by a university to its researchers. This includes RDM services which are built, hosted, and deployed locally, as well as those provided externally but for which the university arranges access for affiliated researchers.

RDM covers a range of complementary yet distinct service categories, with an even wider range of specific services existing within each category. Enumerating all of them does not translate into a checklist of required services a university must deploy in order to build a credible RDM capacity. For example, a university may not feel that offering locally built Curation resources, such as data repository services, is necessary given local circumstances. Nor is the decision necessarily a binary "offer-or-do-not-offer" one; instead, the decision may be one of striking the appropriate emphasis. For example, in deploying Expertise services, a university may choose to establish a general helpline email account to handle RDM-related inquiries, rather than a more elaborate strategy involving data librarians or discipline-specific liaisons providing face-to-face consultation.

Our second report, *Scoping the University RDM Service Bundle*,[4] 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.

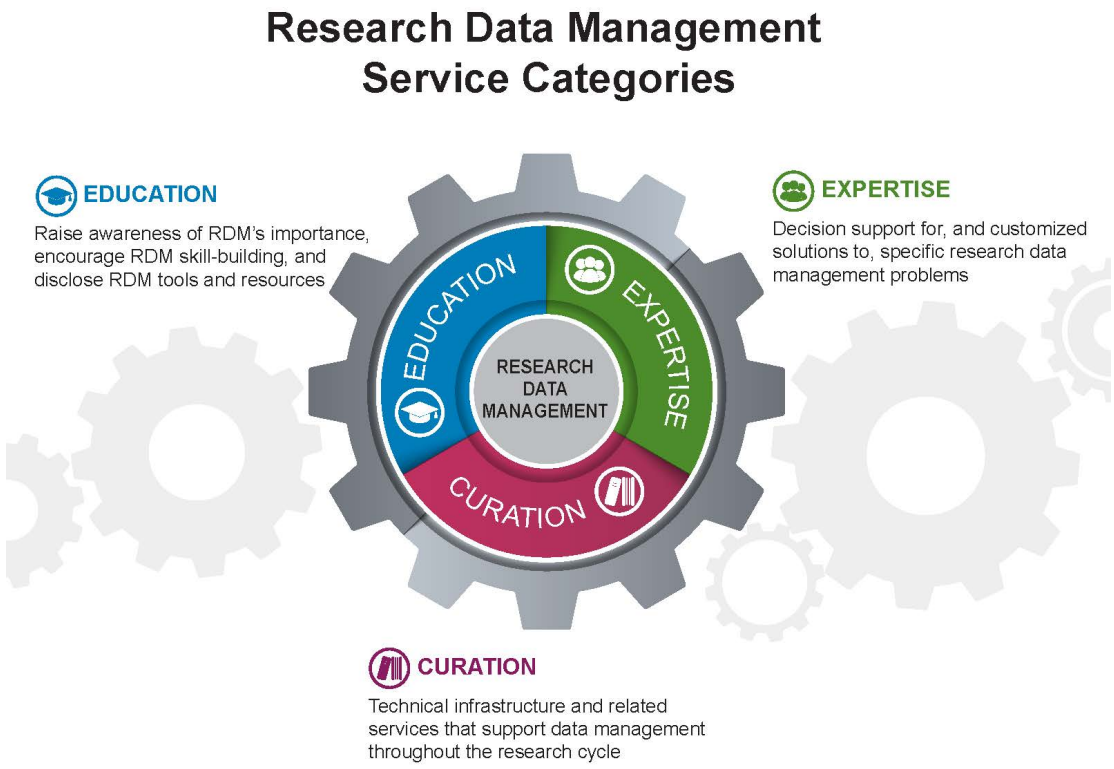


FIGURE 1. RDM SERVICE CATEGORIES[5]

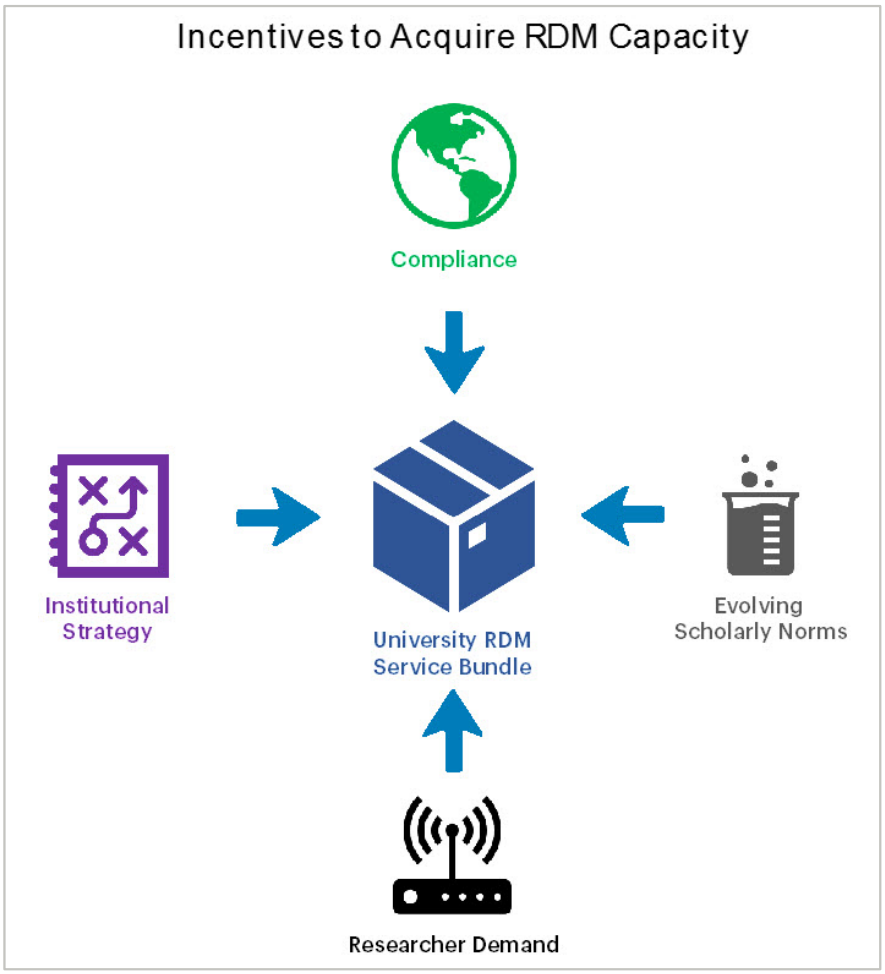


FIGURE 2. INCENTIVES TO ACQUIRE RDM CAPACITY.

As we note in 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 emerging from both local circumstances and the broader environment in which it is situated.

Before choices were made about the services needed to support local RDM requirements, the universities needed an impetus to act—i.e., motivation to allocate resources to RDM services, infrastructure, and other resources. In this report, we focus on the internal and external incentives that prompted each of our case study partners to acquire RDM capacity, or in other words, to develop an RDM service bundle.

Decision Point: Deciding to Act

"Most of economics can be summarized in four words," observes the economist Steven Landsburg. "'People respond to incentives.' The rest is commentary."[6] Incentives are important for understanding any kind of resource allocation, and RDM capacity acquisition is no exception. From an institutional perspective, the first step in addressing RDM is *deciding to act*—in other words, to take steps to meet the RDM needs prevailing at the university.

Incentives or motivations to acquire RDM capacity are multi-faceted and flow from different sources. Based on our case studies, as well as the broader RDM landscape, we organized these incentives into four broad categories, illustrated in figure 2.

Compliance: The promulgation of data management mandates on the part of funders, journals, national agencies, and other external stakeholders is an important external factor influencing university RDM strategy.

Evolving scholarly norms: Increasing interest in data sharing and re-use is part of a broader movement for open science. There is also growing interest in various disciplines on facilitating replication of results as a key element of good scholarly practice, as well as new efforts to reduce the incidence of academic fraud.

Institutional strategy: Many research universities have elevated RDM to an institutional priority—for example, as part of a broader strategy to collect or document the full range of institutional research outputs. Reputation enhancement may be a priority here, and is often linked to local research information management (RIM) system implementations, such as Pure, Elements, or Converis. Universities may also aspire to be a recognized center of excellence in RDM.

Researcher demand: In addition to top-down institutional priorities, RDM capacity acquisition may also be influenced by bottom-up requirements expressed by local/affiliated researchers seeking to close gaps in their scholarly workflows.

Of course, the four categories of incentives described above are not independent or mutually exclusive. For example, changes in scholarly norms regarding data management will influence researchers' perceived workflow requirements, which upon their communication to university decision-makers, may then be prioritized within future institutional strategy around RDM. Nevertheless, these categories are a useful way to summarize the landscape of incentives that research universities are monitoring and responding to. We will refer to these categories of influencing factors throughout the report.

Compliance

Today, national and funder policies and guidelines related to data management are common in all of the national environments we examined. These mandates have largely come about in the last decade and may require data management plans (DMPs) within grant proposals and/or compliance with open data sharing requirements.

RDM activities at three institutions in our study—Edinburgh, Monash, and Wageningen—predate national data curation requirements.

While the first national statements on responsible conduct of research may not have specifically required data management and curation activities, these pronouncements—or advance knowledge that requirements were imminent—may have stimulated the development of *local* policies. For example, the establishment of local policies at Monash in 2010 explicitly followed the mandate described in the 2007 *Australian Code for the Responsible Conduct of Research*, which states, "Each institution must have a policy on the retention of materials and research data. It is important that institutions acknowledge their continuing role in the management of research material and data."[7]

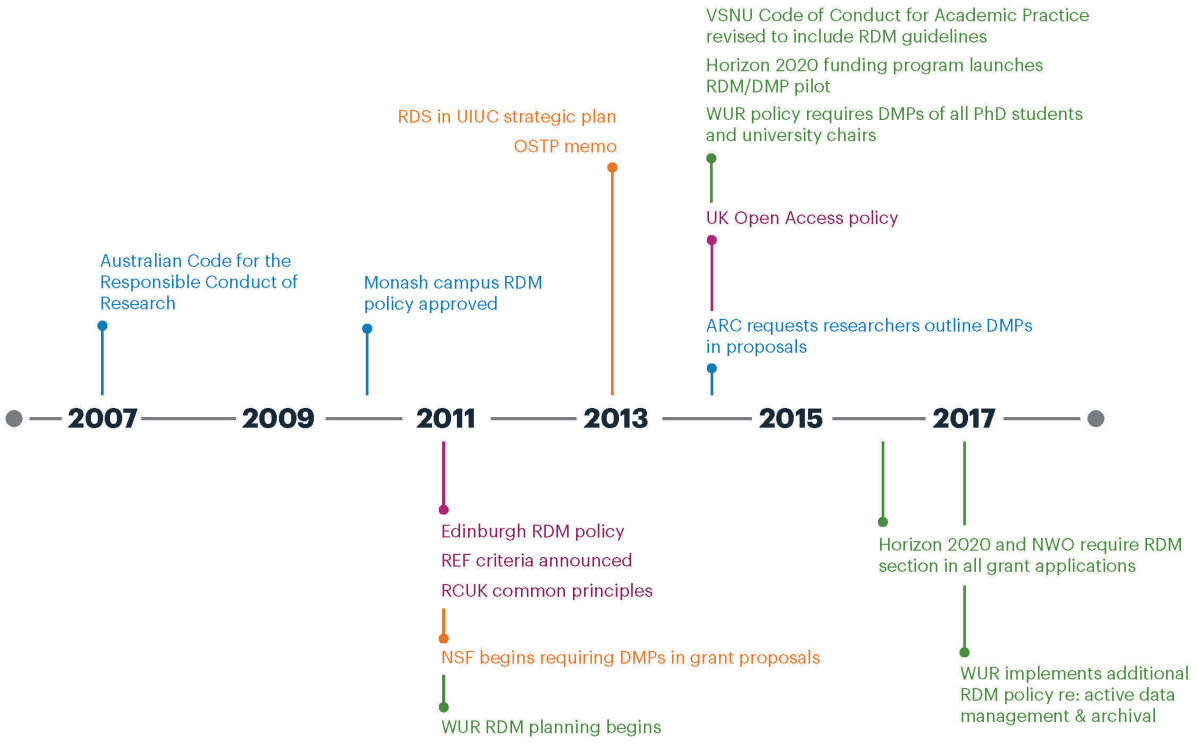


FIGURE 3. TIMELINE OF RDM POLICIES IN FOUR NATIONAL CONTEXTS (MONASH MILESTONES ARE BLUE, EDINBURGH MILESTONES ARE RED, ILLINOIS MILESTONES ARE ORANGE, AND WAGENINGEN MILESTONES ARE GREEN.)

In the UK, the Research Councils UK announced its *Common Principles on Data Policy* in April 2011, stating that "publicly funded research data . . . should be made openly available with as few restrictions as possible in a timely and responsible manner. Institutional . . . data management policies and plans should be in accordance with relevant standards and community best practice."[8] Edinburgh announced its own data management policy only a month later.[9]

It's interesting to note that the RDM activities at three institutions in our study—Edinburgh, Monash, and Wageningen—predate national data curation *requirements*. In our interviews, informants told us how leaders at their institutions anticipated that research data management would be an increasingly important emerging area. For instance, Monash University appointed its first RDM coordinator in 2008, shortly after the 2007 *Australian Code for the Responsible Conduct of Research* statement. Edinburgh and Wageningen had similarly begun planning before national data management requirements were announced.

National policies may define or recommend the length of data preservation, which, in turn, determines institutional practice. For example, VSNU (the association of universities in the Netherlands) has articulated national scientific data retention requirements as part of the Netherlands Code of Conduct for Scientific Practice: research data sets are to be preserved for a minimum of ten years to promote reuse and ensure that scientific work can be validated through replication.[10] The *Australian code for the responsible conduct of research* recommends softer guidelines: "In general, the minimum recommended period for retention of research data is five years from the date of publication. However, in any particular case, the period for which data should be retained should be

determined by the specific type of research."[11] As national policies impacting institutional practice and reporting were unfolding in Europe and Australia, requirements of a different sort were developing in the United States, impacting developments at the University of Illinois and representing one of our key findings: that *differences in external mandates mean that institutions may prioritize different RDM services*. Beginning in January 2011, the National Science Foundation (NSF) began requiring supplemental DMPs in NSF grant proposals. These DMPs were expected to describe how investigators would responsibly manage and share the results and data from NSF-supported research.[12]

Illinois, together with other research institutions highly dependent on NSF extramural support, began work almost immediately to develop the DMPTool, to help researchers comply with this new DMP mandate.[13] Illinois consistently receives more funding from NSF than any other US university,[14] intensifying the local urgency to comply with the NSF DMP requirement. The campus responded by planning and developing comprehensive research data management offerings—that not only directly serve researchers but also enable researchers to demonstrate the local commitment and support for responsible data management by the institution.

Differences in external mandates mean that institutions may prioritize different RDM services.

In February 2013, the Obama administration's White House Office of Science and Technology Policy (OSTP) released a public access memo directing federal agencies supporting research to develop a plan to support increasing public access to publicly supported research. Specifically, the OSTP memo called for the public availability of federally funded research outputs after a 12-month post-publication embargo period. As each federal agency providing research funding was responsible for developing its own public access plan, a complex landscape of differing requirements, timelines, and systems has emerged. The Illinois Research Data Service now has responsibility for monitoring this environment and educating researchers about the specific data sharing requirements of nearly two dozen distinct funding agencies.[15]

In summary, mandates can be an important driver for a university to acquire RDM capacity, but in our case studies, we see this incentive manifesting more as a motivator for *continued* provision of an RDM service bundle, as well as an influencing factor in shaping the type and character of the RDM services included in the service bundle. In this sense, rather than catalyzing a university's acquisition of RDM capacity, mandates play a more prominent role in *shaping or directing* the acquired capacity to conform to national, disciplinary, or funder expectations.

Evolving Scholarly Norms

Institutional provision of RDM services is a byproduct of seismic changes in scholarly practice made possible by rapid technological and network advancements. As articulated in an earlier OCLC Research report, *The Evolving Scholarly Record*, we are witnessing a shift from a traditionally print-centric scholarly record (comprised primarily of text-based materials like journals and monographs), to a more extensive, yet less well-defined scholarly record, that is additionally comprised of materials like datasets, software code, and visualizations.[16]

Scholarly norms are evolving as a result, as researchers, librarians, funders, and policy makers ask how they can responsibly manage this data and facilitate the broadest possible availability in order to foster greater scientific research access, transparency, collaboration,

use, and innovation. Today these values and conversations are coalescing under the umbrella term *open science*, which broadly encompasses numerous components of the research life cycle, including open access to publications, open research data, open source software, open collaboration, open peer review, and open notebooks.

While there is no universally accepted definition of open science, the Organisation for Economic Co-operation and Development (OECD) describes it as: "efforts . . . to make the primary outputs of publicly funded research results—publications and the research data—publicly accessible in digital format with no or minimal restriction." Many professional library organizations have issued statements in support of open science, and, since 2015, the FAIR data principles (Findable, Accessible, Interoperable, Reusable) have received worldwide recognition as a useful framework for thinking about research data management.[17]

Scholarly norms are changing and are enacting pressures on data management practices. But these changes and pressures are not uniform: the evolving practices of one part of the research community may differ from those with another.

The adoption of persistent identifiers and enhanced metadata have arisen as ways to improve citability and appropriate attribution. DOIs have emerged as the most common identifier for datasets, as for journal publications. Some publishers, such as Nature and PLOS, now require submitting authors to share datasets and supporting materials with their article submission, and datasets are published, and given DOIs, in tandem with the published journal article.[18] In response to this emerging scholarly norm of publishing citable data sets, all four of our case study institutions have established processes for minting DOIs as part of the curation workflow.

While there seems to be widespread agreement that data sharing is a good thing, there is not yet widespread agreement on what data sharing *is*. Heidi Imker, University of Illinois at Urbana-Champaign, emphasizes that a single definition of what constitutes data sharing cannot be extrapolated across all disciplines, and there are significant variations between, and even within, domains. And while the data curation community expresses concerns that scientists rarely share research data, there is extensive evidence of data sharing by researchers.

Several surveys and case studies of researcher behavior demonstrate that researchers do share data—although perhaps not in a data repository with well-curated metadata, but upon request. In fact, many scientists think of article publication as a form of data sharing. Scholarly norms *are* changing and are enacting pressures on data management practices. But these changes and pressures are not uniform: the evolving practices of one part of the research community may differ from those with another. This heterogeneity creates significant challenges for data curators seeking to serve researchers across multiple disciplines at comprehensive research institutions.[19]

As the 2017 NMC Horizon Report articulates, research data management is an important emerging trend in research libraries, and we observe institutional pride and prestige among institutions demonstrating domain leadership.[20] Some institutions dedicate library resources to enriching metadata records. For example, Illinois data curators work directly with researchers to

enrich their metadata for optimal discoverability whether depositing in the local Illinois Data Bank or in an external disciplinary repository.

Edinburgh also dedicates considerable effort to enriching metadata and harmonizing information between siloed research information management and data curation systems.[21] Monash provides less direct support for metadata creation, as library curation of datasets is seen as unsustainable. All four of the institutions in our study sought to train researchers on the importance of quality metadata for ensuring discoverability, citability, and reuse through educational programming.

An additional part of the evolving research landscape is the "replicability crisis," in which the results of many scientific studies are difficult or impossible to replicate, bringing the validity of results into question.[22] Open sharing of research data sets is one proposed amelioration to this problem. Open data sharing also brings greater transparency to science, and additional training for scientists on responsible data management may be seen as a way to dissuade academic misconduct, such as the two high-profile research misconduct cases that rocked the Dutch academic community in late 2011. In these cases, two prominent researchers at Dutch universities were fired after their secretive management of fabricated research data was discovered.[23]

The research institutions in our study seek to codify research data management and sharing practices through policy.

We found that the research institutions in our study seek to codify research data management and sharing practices through policy. Edinburgh, Wageningen, and Monash have implemented institutional RDM policies that articulate goals, strategic directions, and specific protocols for researchers. While the Edinburgh policy is openly acknowledged as being "aspirational," it also clearly articulates requirements for data management plans and responsible data management by researchers, as well as the responsibility of the university to "provide training, support, advice and where appropriate guidelines and templates."[24] Wageningen announced its RDM policy in 2014 requiring all PhD students and university chairs to have a DMP, and this policy was expanded in 2017 with additional requirements for responsible active data management as well as long term preservation.[25]

Illinois stands alone among our case study institutions in that it does not have an explicit institutional research data management policy. However, the Illinois Data Service has dedicated significant energy in defining a policy for access and use of the Illinois Data Bank. Through this, Illinois is beginning to tackle the challenging questions of appraising and assessing the enduring value of datasets for long term preservation or deaccession.[26] Through internal collaborations between data curators, university archivists, and digital preservation librarians, it has also explored how libraries can conduct these data preservation assessments at scale.[27] Illinois commits to maintaining each data set for a minimum of five years, after which it may be assessed for continued value as articulated in its preservation review procedures.[28] Edinburgh, on the other hand, intends to exceed the UK requirement to preserve data for ten years and to preserve all locally deposited research data sets indefinitely. These two examples represent significant differences in philosophy and practice, and they also demonstrate what we believe will be important emerging conversations among data curators as research data management practices continue to mature.

Scholarly norms and practices are changing rapidly, and as support for RDM practices, services, and policies grows, this change also

influences and shapes other incentives in our discussion, particularly compliance and institutional strategy.

Institutional Strategy

For a growing number of universities, improving institutional support for research data management is part of a broader strategy to improve process and performance management in the university research enterprise**.** Beyond meeting funder-imposed requirements, universities are motivated to provide better internal tracking of investments in research, as well as their "yield" in terms of publications and related data sets. In national settings, where university funding is allocated from the top down through funding councils and the like, it is increasingly common for universities to license commercial RIM systems that track research productivity on an institutional, departmental, or individual researcher level.

Improving institutional support for research data management is part of a broader strategy to improve process and performance management in the university research enterprise.

Three of the universities in our case studies (Edinburgh, Monash, and Wageningen) register, or plan to register, data sets produced by local faculty and researchers in a local instance of Pure, a RIM system licensed by Elsevier.[29] The University of Illinois has also implemented Pure to assist with producing publicly available faculty research profiles; while the implementation is currently focused on formal publications (journal articles and books), it may be broadened in the future to include data sets managed in the Illinois Data Bank or other repositories.

Monash University is in the process of decommissioning its legacy institutional repository in favor of the in-built repository features of Pure, which provides adequate tracking and reporting functionality to support compliance with funder requirements for open access to research publications. While the university is not required to track research datasets, they regard this as a natural extension of current efforts to track and manage institutional assets.

Wageningen actively encourages faculty and researchers to register datasets in Pure, and provides library support for creating appropriate metadata. Wageningen also monitors external data repositories (disciplinary repositories and commercial repositories like figshare) and creates metadata records in Pure for Wageningen-produced datasets. While they acknowledge that there are gaps in their registry of Wageningen datasets, they are committed to capturing as much metadata for Wageningen research data as possible, because it represents an important part of the total research output of the university.

Like Wageningen, the University of Edinburgh encourages researchers to register their data sets in the local Pure instance; expert data librarians monitor the records to correct and enhance metadata where needed. Creating a comprehensive "university bibliography" is a top priority at Edinburgh, as it assists with internal and external performance management and reporting. Library staff monitor the local DataShare repository to identify data sets that lack a record in Pure and ensure that a new record is created.

University brand management, whether it is upholding an institutional legacy or burnishing an emerging reputation of excellence in a particular discipline, can also be an important driver to acquiring or developing RDM capacity. As early as 2007, Monash recognized that robust support for RDM was necessary to maintaining its prominence as a research

center in data-intensive fields like protein crystallography. The University of Illinois is a large research enterprise in a highly competitive funding environment; maintaining and growing its reputation as a top research university requires local investment in infrastructure and services that will attract top researchers and sponsored research.

According to Heidi Imker, "a key incentive from the institutional perspective was to make it clear in grant applications that Illinois has robust capacity to support funded projects."

This incentive emerged long before the OSTP mandate was in place. In 2013, the office of the provost and officer of chancellor initiated a campus-wide review of IT that revealed an important gap in services and infrastructure for research data management. The University of Illinois is a highly productive research institution producing many research publications and associated data sets each year, but was (at the time of the cyberinfrastructure review) doing little to support research data management.[30]

Developing RDM expertise is key to a strategic agenda to increase library support for research workflows, shifting attention away from traditional back-office activities.

Edinburgh sees itself as a pioneer in developing research data management solutions, and even a potential service provider to the nationally shared Research Data Service currently under development by Jisc. While the general trend within the university has shifted from a "build" to a "buy" approach for IT infrastructure, locally developed solutions for Research Data Management are still preferred, as they contribute to institutional differentiation and help to burnish the university's brand. Additionally, as an "ancient university," Edinburgh feels a particular pride in its role as a steward of the scholarly record. As Dominic Tate, Edinburgh University Library's Scholarly Communications Manager, put it, "From a preservation point of view, [the library] intends to manage [research data sets and other scholarly materials] in perpetuity, as we have been doing since 1583."

Wageningen is developing a Data Competence Center to accelerate the university's "big data" research and education agenda; excellence in RDM service is part of a broader institutional strategy to increase the university's national and international research profile across multiple disciplines (animal science, plant sciences, sustainable development).[31]

For some university libraries, acquiring local RDM capacity can advance a strategic interest in shifting the focus of library service provision from traditional collection-centric functions toward more direct engagement in supporting individual researchers and advancing the institutional research enterprise. At Monash University, developing RDM expertise is key to a strategic agenda to increase library support for research workflows, shifting attention away from traditional back-office activities focused on content acquisition, licensing, or cataloging toward services that support content creation.[32]

David Groenewegen observed: "I think we'll see some shift away from the cataloguing of stuff we got from somewhere else to describing the stuff we have created locally. [Research support is] a tricky area to get into, to be credible in, to continue to grow skills in. How do we go beyond saying 'these are the things you need to know about' to providing researchers with support in their own workflows? It will mean giving up some other things [in the library service portfolio]" in order to support more engagement around research support.

Researcher Demand

Basic economic intuition suggests that a key motivator for developing an RDM service bundle would be an expression of need on the part of a university's researchers. In other words, universities supply RDM services because their researchers demand them. Incentives of this kind are represented in figure 2 by the category *Researcher Demand*: bottom-up demand for RDM services and infrastructure articulated by a university's affiliated researchers to fill perceived gaps in their scholarly workflows.

While clear signals of strong demand from prospective users are indeed an obvious reason for a university to move to acquire RDM capacity, they appear to play a relatively minor role in our four case study partners' decision to act: none of them cited bottom-up demand from affiliated researchers as a primary motivation for developing a local RDM service bundle, and, indeed, several acknowledged that demand for RDM capacity remains low even following deployment of the RDM service bundle on campus. In short, the decision to act—to acquire RDM capacity—was for the most part a response to factors other than demand from local researchers. RDM service bundles were developed (at least initially) in the absence of strong demand from their intended users. Rather than responding to researcher demand, the decision to act tended to *anticipate* this demand.

Monash University provides an instructive example. Acknowledging that the impetus for developing its RDM service bundle did not originate with "faculty clamoring at the door," and that campus demand for RDM services still had not reached a tipping point, Monash RDM staff prioritize engagement with researchers to educate them on the importance of good data management practices, while at the same time pointing to the availability of RDM solutions to back them up. The latter point is especially noteworthy: Monash believes that when researchers are ready to seek RDM support from the university, it is important to have a mature offering for them. The existence of concrete RDM services helps to elucidate, and eventually strengthen, the perceived benefits of data management.

Like Monash, Illinois' RDM service bundle—in particular, the Illinois Data Bank—was developed in response to factors other than local researcher demand. Illinois staff note that the Data Bank remains a "bleeding edge" service, with demand yet to reach its full potential. Wageningen also characterized the impetus behind its RDM service bundle as primarily top-down, catalyzed by a Graduate School policy proposal, which took effect in 2014, requiring doctoral students and research groups to produce data management plans for their research projects.

Researcher demand can be far more important in shaping and sustaining RDM service bundles, rather than incentivizing their creation.

Although researcher demand was not cited as the primary incentive for its RDM service bundle, Edinburgh did note a long-time institutional interest in data management, stemming from the university's strong reputation in informatics and the establishment of the Data Library—a service to assist researchers in the discovery, use, and management of data—in the 1990s. This is believed to have fostered a stronger culture of understanding of the benefits from RDM than might be found at other institutions.

While not driving the *initial* development of RDM service bundles, demand by researchers is nevertheless a matter of keen interest to RDM staff and is viewed as something to be cultivated and strengthened over time. In this sense, researcher demand can be far more important in *shaping and sustaining* RDM service bundles, rather than incentivizing their creation.

Because researcher demand for RDM services are expected to develop and evolve over time, Education services (see figure 1) have emerged as an important component of the RDM service bundles of our case study partners. As mentioned above, Monash prioritizes outreach services and capability building aimed at educating researchers on the importance and benefits of good data management, as well as the RDM solutions Monash offers. For example, a data librarian emailed two thousand Monash researchers, and spoke to several hundred of them, in an effort to identify data sets for curation. There is an evangelistic aspect to this outreach—although researchers may not be ready to deposit data at the time the outreach occurs, they are at least made aware that data management services are in place to support them when needed.

Active data management is one aspect of the RDM service space where bottom-up demand by researchers is particularly relevant.

Similarly, Illinois emphasizes outreach to researchers as an important element of its Education services, and, indeed, staff estimate that activities such as workshops and training are the most-used aspects of the RDM service bundle. An important goal of Illinois's researcher outreach is correcting the misperception that data management is just about *ex post* data sharing; therefore, emphasis is placed on educating researchers about good active data management practices. Edinburgh also deploys an extensive outreach program, with an emphasis on engaging research administrators—a "train the trainers" approach—via internal mailing lists and wikis.

Successful outreach programs can, over time, cultivate the demand that will help establish RDM as a critical piece of scholarly infrastructure, as well as bolster the case for ongoing funding and support for the RDM service bundle. Furthermore, outreach programs can play an important role in helping RDM staff stay abreast of fluid researcher data management needs. For example, Wageningen's RDM service bundle originally developed around DMP planning, in response to an institutional initiative requiring DMPs from Wageningen researchers.

However, as the service bundle continued to evolve, new services were added, including a GitLab repository implementation to manage source code storage and sharing. Wageningen RDM staff note that this was done in response to demand from researchers—something researchers indicated they wanted and needed. The Illinois RDM service bundle also evolved in response to emerging demand: the recently launched active data management service at Illinois addresses an identified need for rentable mid-range data storage.

Illinois's experience with active data management exemplifies a broader pattern in how RDM service bundles are being scoped and evolved in response to emerging demand. Like Illinois, Wageningen has also expanded its RDM service bundle to include more support for active data management, which, as Wageningen RDM staff point out, was driven from researcher demand. DataStore, Edinburgh's active data management solution, receives the most usage among the university's array of Curation services, while Monash also offers a variety of active data management services to its researchers.

Indeed, active data management is one aspect of the RDM service space where bottom-up demand by researchers is particularly relevant. While RDM may have been originally cast as a long-term data curation problem, with a focus on preservation of data at the end of the research process, it may be that demand from researchers have altered that view, shifting it to

encompass data management at all points in the research lifecycle, and instigating a corresponding shift in the scoping of the RDM service bundle.

Researcher demand is particularly important as a source of intelligence in evolving the RDM service bundle given ongoing uncertainty about the future of RDM, both as a set of accepted scholarly practices and as an ecosystem of services and infrastructure. For example, Edinburgh staff observed that looking ahead

even within a modest time frame—three to five years—is difficult: researcher RDM needs are still unclear, and the situation is fluid. This underlines the importance of continuously monitoring local demand as a routine part of managing the RDM service bundle. Researcher engagement—talking directly to those whom the service bundle is intended to support—is essential. Illinois RDM staff note that researcher engagement is reported as part of the RDM service bundle's performance metrics, and is an important element in demonstrating to university administrators the value of sustaining RDM capacity.

A key finding from our earlier report, *Scoping the University RDM Service Bundle*, is that RDM is not a monolithic service, duplicated from university to university, but is instead a customized solution tailored to the specific needs and circumstances of each institution. In the same way, *demand* for RDM services is not monolithic either: in particular, the demand from researchers will tend to differ across institutions. Certain RDM services may enjoy higher expressions of demand at some universities than at others. For example, RDM workshops and training are the services that enjoy the highest usage at Illinois; in contrast, Edinburgh reports that DataStore—a file store for active data management—receives the most usage.

But perhaps even more important than cross-institutional differences in demand are differences that manifest *within* institutions. Demand for RDM services can vary significantly from discipline to discipline. Some academic units may be quite keen to utilize local RDM capacity and may even catalyze a broader university RDM strategy. Monash, for example, cites early engagement with researchers in protein crystallography as an important precursor to the development of an RDM service bundle, exemplifying a campus cohort with a concrete RDM need.

But cross-disciplinary differences in RDM demand patterns can work in the other direction as well. For example, some disciplines may be interested in availing themselves of RDM capacity, but seek solutions external to the university. Edinburgh staff observed that physicists tended to use CERN resources for their RDM needs, rather than local solutions. The experiences at Monash and Edinburgh are suggestive of a broader lesson: monitoring demand requires digging deeper than "overall" demand metrics for RDM services; RDM staff should seek to uncover more granular demand patterns at the disciplinary level.

It is important to bear in mind that researcher demand for RDM services is usually a *derived demand*. The goal is not to manage data for the sake of managing data; rather, the demand for RDM services is derived from demand for the myriad benefits potentially realized from good data management practices: e.g., reputation enhancement, availability of data for new research, replication of findings, improved collaboration, and so on. Given this, the ability of RDM staff to strengthen researcher demand for RDM services will depend on making appropriate connections between RDM services and the end-benefits that are of most interest to researchers. For example, Edinburgh staff found that emphasizing a positive message of the *benefits* of data sharing was more effective in capturing the interest of researchers than highlighting the *problem* (and administrative burden) of compliance with funder requirements. Strengthening and responding effectively to demand requires a thorough understanding of the perceived benefits motivating demand.

Conclusion

In this report, we examined some of the key incentives that motivated the universities profiled in our case studies to acquire or develop institutional capacity to support research data management. While specific motivations varied from one institutional context to the next, we identified four general categories of incentives (see figure 2) that influence the "decision to act" in each of the four universities we studied:

1. **Compliance** with mandates or policies that establish formal requirements for documenting research data management plans or for demonstrating progress toward open science goals.

2. **Evolving scholarly norms** that influence disciplinary perspectives on what constitutes good scientific practice, including expectations of reproducibility and transparency in documenting protocols, methods, and data sources.

3. **Institutional strategies** that are aided by more rigorous and systematic attention to monitoring research productivity and performance, and improving (or maintaining) institutional reputation in data-intensive research areas.

4. Direct or derived **demand** from researchers with unmet (or imperfectly satisfied) data management needs; for example, evidence that university researchers are turning to external services to meet data storage, management, or sharing needs that could be met by the university.

A key takeaway from our investigation is that **university investment in research data management infrastructure, services, or personnel is motivated by *locally relevant incentives*.** In other words, the increased attention to RDM in research universities operating in different local circumstances reflects an alignment of institutional interests (to maximize grant funding, burnish research reputation, or leverage distinctive capacities) and external motivations (policy mandates, scientific norms, and evolving research workflows).

Another important takeaway is that **our case study partners acted to establish RDM services *in anticipation of*, rather than in direct response to, researcher demand and explicit policy mandates.** Incentives related to institutional strategy and evolving scholarly norms played a larger role in directly catalyzing RDM service development at these institutions. Researcher demand and compliance with policy mandates were important factors in re-shaping and sustaining the RDM service bundle over time, but were not the key drivers for establishing RDM services in our case study institutions.

While the constellation of relevant incentives differs from one context to another, the acquisition or development of local RDM capacity is invariably motivated by an interest in protecting or enhancing institutional reputation and success. Consequently, the long-term sustainability of university RDM services is contingent upon alignment with institutional needs, as much as individual researcher needs. Put another way, RDM is not merely a fad but instead represents a rational institutional response to powerful, if transitory, incentives.

ACKNOWLEDGMENTS

The authors extend special thanks to our interview informants: Heidi Imker and Beth Sandore Namachchivaya (University of Illinois at Urbana-Champaign); David Groenewegen (Monash University); Dominic Tate and Jeremy Upton (University of Edinburgh); and Jacquelijn Ringersma and Ellen Fest (Wageningen University & Research).

We note with appreciation that the libraries participating in this project include three members of the OCLC Research Library Partnership (Illinois, Monash, and Edinburgh), and the home institution of an OCLC Europe, the Middle East and Africa (EMEA) Regional Council member (Hubert Krekels, Library Director at Wageningen).

Thanks are also due to Lorcan Dempsey, Vice President, Membership and Research, Chief Strategist, for his role encouraging a program of work in the area of research data management, to previous work done in the area of RDM by OCLC Research colleagues Ricky Erway and Ixchel Faniel, and to our OCLC colleagues Erin M. Schadt, Jeanette McNicol, and JD Shipengrover for their help in publishing this report.

NOTES

1. 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, OH: OCLC Research. doi:10.25333/C3PG8J.
2. 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. doi:10.25333/C3Z039.
3. See note 1.
4. See note 2.
5. See note 1.
6. Landsburg, Steven. 2012. *The Armchair Economist: Economics and Everyday Life.* 2nd Ed. New York: Free Press, 3.
7. Australian Government. 2007. *Australian Code for the Responsible Conduct of Research.* Australian Research Council. http://www.nhmrc.gov.au/_files_nhmrc/file/publications/synopses/r39.pdf.
8. Research Councils UK. "RCUK Common Principles on Data Policy." http://www.rcuk.ac.uk/research/datapolicy/.
9. https://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy.
10. The Hague. 2004.*The Netherlands Code of Conduct for Academic Practice Principles of Good Academic Teaching and Research.* Translated by Metamorfose Vertalingen BV. Den Haag: Association of Universities in the Netherlands. http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/The_Netherlands_Code%20of_Conduct_for_Academic_Practice_2004_(version2014).pdf.
11. See note 6.
12. https://www.nsf.gov/eng/general/dmp.jsp.
13. https://dmptool.org/about. [DMPTool was released in October 2017.]
14. https://dellweb.bfa.nsf.gov/Top50Inst2/default.asp.
15. https://www.library.illinois.edu/scp/ostp-mandate/; https://web.archive.org/web/20161222160413/https://www.whitehouse.gov/administration/eop/ostp/library/publicaccesspolicy; Imker, Heidi J., and Daniel G. Tracy. 2015. "Public Access to Articles and Data for Funding Agencies." Presentation given at University of Illinois at Urbana-Champaign. http://hdl.handle.net/2142/79494.
16. Lavoie, Brian, Eric Childress, Ricky Erway, Ixchel Faniel, Constance Malpas, Jennifer Schaffner, and Titia van der Werf. 2014. *The Evolving Scholarly Record. Dublin, Ohio: OCLC Research*. http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-evolving-scholarly-record-2014.pdf.

17. OECD (Organisation for Economic Co-operation and Development). 2015. *Making Open Science a Reality*. OECD Science, Technology and Industry Policy Papers, no. 25. Paris: OECD Publishing, 7. doi:10.1787/5jrs2f963zs1-en; For example, see the support for open science by LIBER, http://libereurope.eu/strategy/strategic-direction-1-enable-open-science/; https://www.nature.com/articles/sdata201618.

18. http://www.nature.com/authors/policies/availability.html?foxtrotcallback=true#data; Since 2014, "PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception." http://journals.plos.org/plosone/s/data-availability.

19. Imker, Heidi J. 2017. "Overlooked and Overrated Data Sharing: Why Some Scientists are Confused and/or Dismissive." In *Curating Research Data, Volume One: Practical Strategies for Your Digital Repository.* edited by Lisa R. Johnston, 127-150. Chicago, IL: Association of College and Research Libraries. http://hdl.handle.net/2142/95024.

20. Adams Becker, S., M. Cummins, A. Davis, A. Freeman, C. Giesinger Hall, V. Ananthanarayanan, K. Langley, and N. Wolfson. 2017. *NMC Horizon Report: 2017 Library Edition.* Austin, Texas: The New Media Consortium. http://cdn.nmc.org/media/2017-nmc-horizon-report-library-EN.pdf.

21. At Edinburgh, researchers are expected to deposit research dataset metadata into the Pure research information management system, in order to comply with the UK Open Access policy. Data librarians monitor incoming records to ensure quality metadata for reporting. They also work to ensure that any datasets deposited directly into the local Datashare repository are consistently registered in Pure. Currently the two systems do not interoperate, so processes are cumbersome and duplicative.

22. Baker, Monya. 2016. "Is There a Reproducibility Crisis?" *Nature* 533: 452–54. doi:10.1038/533452a.

23. "Breaking News: Prolific Dutch Heart Researcher Fire Over Misconduct Concerns." 2011. *Retraction Watch* (blog). Posted 17 November. http://retractionwatch.com/2011/11/17/breaking-news-prolific-dutch-heart-researcher-fired-over-misconduct-concerns/; Verfaellie, Mieke, and Jenna McGwin. 2011. "The Case of Diederik Stapel: Allegations of Scientific Fraud by Prominent Dutch Social Psychologist are Investigated by Multiple Universities. *Psychological Science Agenda*. Posted December. http://www.apa.org/science/about/psa/2011/12/diederik-stapel.aspx; Callaway, Ewen. 2011. "Report Finds Massive Fraud at Dutch Universities: Investigation Claims Dozens Of Social-Psychology Papers Contain Faked Data." *Nature*. 479: 15. doi:10.1038/479015a; In addition to fabricating data, secretive practices included undisclosed survey instruments and results, doctoral students who did not collect their own data, and data that was unavailable for later use.

24. University of Edinburgh. 2011. "Research Data Management Policy." Posted 16 May. https://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy.

25. OCLC Research hosted a webinar entitled, "Policy Realities in Research Data Management," in which Jacquelijn Ringersma, Head of Digital Production Centre, Wageningen University & Research and Heidi Imker, Director of the Research Data Service (RDS) at the University of Illinois at Urbana-Champaign spoke about policy developments at their institutions. Their slides and a recording of their presentations are available at http://www.oclc.org/research/events/2017/07-25.html.

26. Illinois does have a general Policy and Procedures on Integrity in Research and Publication, and while it explicitly defines research misconduct as including the fabrication or falsification of data, it doesn't explicitly require responsible conduct of research and data management best practices or plans from researchers. See *Policy and Procedures on Integrity in Research and Publication*. 2009. https://www.vpaa.uillinois.edu/UserFiles/Servers/Server_420372/File/Integrity-Policy.pdf and https://databank.illinois.edu/policies; Heidi Imker spoke about how Illinois worked to develop these policies in the 2017 OCLC Research webinar, "Policy Realities in Research Data Management." http://www.oclc.org/research/events/2017/07-25.html.

27. Anderson, B., S. Braxton, E. Dunham, H. Imker, and K. Rimkus. 2016. "Should We Keep Everything Forever? Determining the Long-Term Value of Research Data." Poster presented at iPRES 2016, 13th International Conference on Digital Preservation, 3-6 October 2016, Bern, Switzerland. http://hdl.handle.net/2142/91659.

28. "Preservation Review, Retention, Deaccession, Revision, and Withdrawal Procedure." Illinois Data Bank, University of Illinois at Urbana-Champaign. Accessed 6 November 2017. https://databank.illinois.edu/policies#preservation_review.

29. For additional examples of universities using RIM systems to register and manage research data sets, see Schöpfel, Joachim, Hélène Prost, and Violane Rebouillat. 2017. "Research Data in Current Research Information Systems." *Procedia Computer Science*, 106:305–20. doi:10.1016/j.procs.2017.03.030 and Clements, Anna, and Valerie McCutcheon. 2014. "Research Data Meets Research Information Management: Two Case Studies Using (a) Pure CERIF-CRIS and (B) EPrints Repository Platform with CERIF Extensions." *Procedia Computer Science* 33: 199–206. doi:10.1016/j.procs.2014.06.033.

30. Towns, John, David Gerstenecker, Laura Herriott, Ashley Hetrick, Heidi J. Imker, Chris Larrison, Marshall Scott Poole, Eric Gene Shaffer, Tracy Smith, Chuck Thompson, and Rob Watson. 2015. *University of Illinois Year of Cyberinfrastructure Final Report.* Champaign, IL: University of Illinois at Urbana-Champaign. https://www.ideals.illinois.edu/handle/2142/88444.

31. http://www.wur.nl/nl/Dossiers/dossier/Big-Data.htm.

32. While only one of the institutions profiled in this report explicitly mentioned RDM as being part of a strategic shift from collection-centric operations to a stronger focus on researcher support, other OCLC research has identified this as an emerging trend in academic libraries. See, for example: Dempsey, Lorcan. 2016. "Library Collections in the Life of the User: Two Directions." *LIBER Quarterly* 26(4): 338-359. doi:10.18352/lq.10170; 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: http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-collectiondirections-preprint-2014.pdf.

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

6565 Kilgour Place
Dublin, Ohio 43017-3395

T: 1-800-848-5878
T: +1-614-764-6000
F: +1-614-764-6096
www.oclc.org/research

ISBN: 978-1-55653-043-2
DOI: 10.25333/C3S62F
RM-PR-215833-WWAE-R3 1712