NOVEMBER 2021

OCLC RESEARCH REPORT

PART ONE

Findings and Recommendations

Research Information Management in the United States

OCLC

Research Information Management in the United States: Part 1—Findings and Recommendations

Rebecca Bryant
OCLC Research

Jan Fransen
University of Minnesota

Pablo de Castro
Strathclyde University and euroCRIS

Brenna Helmstutler
Syracuse University

David Scherer
Carnegie Mellon University

October 2021

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

ISBN: 978-1-55653-224-5
DOI: 10.25333/8hgy-s428
OCLC Control Number: 1281710894

ORCID iDs
Rebecca Bryant, OCLC Research https://orcid.org/0000-0002-2753-3881

Jan Fransen, University of Minnesota https://orcid.org/0000-0002-0302-2761

Pablo de Castro, Strathclyde University and euroCRIS https://orcid.org/0000-0001-6300-1033

Brenna Helmstutler, Syracuse University https://orcid.org/0000-0001-5549-9935

David Scherer, Carnegie Mellon University https://orcid.org/0000-0002-6244-4331

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

Suggested citation:
Bryant, Rebecca, Jan Fransen, Pablo de Castro. Brenna Helmstutler, and David Scherer. 2021. *Research Information Management in the United States: Part 1—Findings and Recommendations*. Dublin, OH: OCLC Research. https://doi.org/10.25333/8hgy-s428.

CONTENTS

TABLES AND FIGURES

INTRODUCTION

Research information management (RIM) is a rapidly growing area of investment in US research universities. RIM systems that support the collection and use of research outputs metadata have been in place for many years. Globally, the RIM ecosystem is quite mature in locales where national research assessment exercises like the United Kingdom's Research Excellence Framework (REF) and the Excellence in Research for Australia (ERA) require institutions to collect and report on the outputs of institutional research. A pan-European community of practice is led by euroCRIS.[1]

However, in the absence of national research assessment requirements in the United States, RIM practices at US research universities have taken a different—and characteristically decentralized—course. A complex variety of stakeholders have responded to a similarly complex mix of use cases, with silos and frequent duplication as a result.

The *Research Information Management in the United States* two-part report series provides a first-of-its-kind documentation of RIM practices at US research universities, building on previous research conducted by OCLC Research,[2] that offers a thorough examination of RIM practices, goals, stakeholders, and system components. Part 1 provides summary findings synthesized from in-depth case studies of US RIM practices, which are documented in detail in the Part 2 companion report. It furthermore synthesizes these findings into a summary of RIM use cases, a RIM system framework, and concise recommendations for RIM stakeholders, providing much-needed context for institutional leaders to examine their own local practices.

The *Research Information Management in the United States* two-part report series provides a first-of-its-kind documentation of RIM practices at US research universities.

This report describes six discrete RIM use cases detailed in the companion report:

- Faculty Activity Reporting (FAR)
- Public Portal
- Strategic Reporting and Decision Support
- Metadata Reuse
- Open Access Workflow
- Compliance Monitoring

These use case scenarios describe the functional goals and business needs that institutions may be seeking to satisfy. These business goals are often pursued in silos by different campus stakeholders but share many similarities in data collected and used. Our hope is that our identification of six discrete use cases will provide a frame of reference for institutions to examine and better understand their own complex practices, inviting increased collaboration, information sharing, decision-making, and institutional investment. It also informs our recommendations for institutional leaders, as we advise investments in institutional data curation, widespread adoption of persistent identifiers, and a cross-functional, enterprise approach to institutional research information management.

In addition, this report seeks to establish a unified definition of research information management that embraces the disparate and siloed uses so prevalent in the United States, a necessary first step toward the development of a cross-functional, collaborative, and vendor-agnostic community of practice in the United States. This unified definition is also intended to help a broad array of institutional stakeholders adopt an enterprise view of RIM practices—examining silos, redundancies, duplication of effort, and providing insights into opportunities for improved interoperability, decision support capabilities, and informed institutional investment.

This unified definition is also intended to help a broad array of institutional stakeholders adopt an enterprise view of RIM practices.

The report introduces a RIM system framework, which visualizes the functional and technical components that make up an institution's specific RIM system through subdivision of RIM processes into three discrete segments:

- Data sources
- Data processing (including storage)
- Data consumers

Part 1 concludes with a list of six high-level recommendations for US institutional leaders and decision makers to:

- Invest in institutional data curation
- Support adoption of persistent identifiers (PIDs)
- Don't expect a turn-key system
- Support cross-functional teams
- Invest in dedicated personnel
- Include research information in enterprise data governance efforts

We developed this framework to consistently and comparatively describe the RIM systems documented in Part 2 of this report series, which describes the RIM practices of five US research institutions:

- Penn State University
- Texas A&M University
- Virginia Tech
- UCLA
- University of Miami

The detailed narratives in *Research Information Management in the United States: Part 2—Case Studies* provide the source material and evidence for this report and provide an opportunity for readers to learn more about the RIM history, goals, use cases, scope, stakeholders, and users at each case study institution. We expect readers to dip in and out of these as they identify points of interest.

WHO SHOULD READ THIS REPORT?

This report is written with three overlapping audiences in mind.

Institutional leaders at research universities

First, we have written this report for senior leaders at research universities, including provosts and vice presidents of research. This report exposes the enterprise nature of RIM practices and provides institutional decision makers with both a categorization of uses and in-depth examples that can aid in the understanding of local practices and opportunities. RIM systems offer the potential to support an array of institutional needs such as reputation management and annual academic reviews, as well as strategic planning and decision-making. Read the Recommendations for institutional leaders.

Research library leaders

Our second audience is research library leaders, as the library is an essential stakeholder in enterprise RIM practices, contributing expertise in metadata, scholarly communications, and bibliometrics that are essential for successful RIM implementations. This report documents the extensive role of libraries in RIM and advocates for a role for libraries in all RIM implementations to ensure that publications metadata is rich and fit for purpose.

RIM practitioners

Third, we have written this report to be useful for RIM practitioners who will benefit most from the plentiful details in the Part 2 case studies. We hope that by exposing the enterprise nature of RIM activities, RIM practitioners will see new opportunities for collaboration within their institutions. This report documents how some project details, like implementing the institutional hierarchy in the RIM system, are universally difficult, which may help practitioners explain these challenges to others in their institutions.

SCOPE AND METHODS

This study is focused on RIM practices in US research universities. Because of the decentralized development of RIM activities with multiple use cases, the lack of a centralized community of practice, and confusion about just what comprises RIM practice, there is little US literature that addresses this landscape in a holistic way. The existing literature largely reflects silos of practice, with publications on specific implementations, products, and uses.

The 2018 report *Practices and Patterns in Research Information Management: Findings from a Global Survey* produced by OCLC Research and euroCRIS attempted to address this gap.[3] However, the survey effort yielded a small and unrepresentative sample of US institutions. As current and former RIM practitioners, we knew that there was a great deal of activity in the United States, but that a global survey was insufficient for capturing information about the complex heterogeneity of stakeholders and business uses specific to this region. Instead, we concluded that in-depth case studies of a few institutions would offer a more nuanced perspective of the RIM ecosystem in the United States, and we selected five institutions to examine:

- Penn State University
- Texas A&M University
- Virginia Tech
- UCLA (including the University of California system-wide practices)
- University of Miami

The institutions selected for study were chosen because they represented a diversity of:

- Use cases
- Stakeholders
- Products (including proprietary, open source, and homegrown solutions)
- Scale (encompassing both institutional practices and those occurring at the system or state level)

In examining RIM practices, we sought to document efforts that aggregated information about research outputs, particularly publications.[4] This led us to examine public portal use cases of products like VIVO and Pure, as well as faculty activity reporting (FAR) systems that aggregated similar information to support annual academic progress reviews and/or promotion and tenure (P&T) processes. It also led us to explore the connection between a RIM system aggregating an institutional bibliography and the institutional repository that aggregates and shares the scholarly content.

If a university maintained a faculty activity reporting system for all (or most) of the institution, we included this in our study, leading us to document the practices at Penn State, Virginia Tech, and UCLA. However, we did not document FAR practices at Texas A&M and the University of Miami because the FAR practices there are currently decentralized, managed by multiple colleges using an array of systems. This endeavor would have taken considerable effort and made for a lengthier report, likely without corresponding benefits.

Data was collected through semi-structured interviews with individuals supporting or using RIM infrastructures at the case study institutions in an array of roles and units, including the library, research administration, information and communications technology, and academic affairs.

Prior to each interview, participants received general information about the project as well as the interview protocol (see appendix available in *Research Information Management in the United States: Part 2—Case Studies*).[5] All interviews were conducted via videoconference technology with at least two researchers participating, and they were also recorded for later review exclusively by research team members. In total, we interviewed 39 individuals from eight institutions in 23 separate interviews.[6] Because of the nature of this study, we were unable to assure anonymity. We have identified participants in the Acknowledgments section of this report but do not attribute any specific comments to any specific person. We also use the non-gendered pronoun “they” when referring to our interview participants.

Information gathered from these interviews was combined with a review of the existing scholarly literature and significant gray literature in the form of system documentation, white papers, and more. Each interview participant was invited to review a draft of the case study about their institution to support accuracy.

LIMITATIONS

This study provides a detailed overview of RIM practices at five US institutions. However, it does not purport to describe all RIM practices in the United States nor to provide an example of “typical” RIM practices. While US research universities share many similarities, RIM practices are far from standardized. Each institutional story will be somewhat unique.

While we’ve aspired to collect the same information about each system, it wasn’t always possible to secure. For instance, RIM administrators at one institution found it difficult to tell us how many person profiles were included in their system while another found it impossible to estimate the aggregate FTE supporting their project. We have sought to be as detailed and accurate as possible, including the relevant information collected.

This study is also not intended as a comparison of different RIM products. Some comparative information is available through the case studies in Part 2, but it is not an intentional deliverable of this report. Products in this landscape are evolving rapidly, and it is our view that any comparative analysis of products would be immediately obsolete. Also note that this study documents how institutions are *using* a product, which might not include the full range of functionality a product might support. Furthermore, these case studies are a snapshot of institutional practices in 2021, and we acknowledge that institutional practices will continue to evolve and change.

With these challenges in mind, we opted to assemble and share the RIM adoption stories of five research universities, allowing us to compare and contrast their activities, synthesize common use cases, and provide generalized insights.

The RIM landscape in the United States

US RIM SYSTEMS OCCUR AT MANY SCALES

This study documents the scale of institutional practices, examining RIM systems deployed at the college, campus, system, or even state levels. The development of RIM infrastructure at any institution reflects local practices, histories, and individuals, resulting in institutionally unique practices, which is unsurprising given the decentralized and highly autonomous nature of US research universities. A few key observations from this project include:

- Most institutions have multiple RIM systems.
- RIM systems that support faculty activity reporting are currently decentralized at two of the five case study institutions (Miami and Texas A&M).
- RIM systems may exist at the college, campus, and system level in a single institution (UCLA).

- Tenure-line faculty were included in all of the systems documented in this study. Some systems have expanded to also include other researchers, such as nonfaculty researchers (Penn State Research Portal) and graduate students and postdocs (Virginia Tech).

It seems likely that the US RIM systems will follow the lead of institutions like Penn State and Virginia Tech and include more nonfaculty researchers. This inclusion can support a more comprehensive view of institutional research activity and further showcase the activities of research units and graduate programs.

DEFINING RIM SYSTEMS: WHAT DO WE CALL THIS THING?

There is a lack of unified understanding in the United States about just what constitutes a research information management system. This is apparent through both unstandardized nomenclature as well as multiple definitions of RIM system practices. As recognized by Schöpfel and Azeroual, "the nomenclature for research information systems is more or less unstandardized,"[7] but includes terms such as:

- Research Information System (RIS)
- Current Research Information System (CRIS)
- Research Information Management System (RIMS)
- Expert Finder System (EFS)
- Research Networking System (RNS)
- Research Profiling System (RPS)
- Faculty Activity Reporting (FAR) system

The lack of standardized terminology is particularly acute in the United States, which we see as a direct result of the diffuse uses giving rise to different systems supported and used by disparate stakeholders.

In contrast to the European environment where the term Current Research Information System (CRIS) is widely used and understood to "stand for a family of information systems (databases, decision support systems, and so forth) to store and manage data about research conducted at an institution,"[8] there is little consensus in the United States about what constitutes a research information management system.

The lack of standardized terminology is particularly acute in the United States, which we see as a direct result of the diffuse uses giving rise to different systems supported and used by disparate stakeholders.

Some descriptions of RIM systems are largely consistent with European notions of what a CRIS system is, such as "RIMS collect and store structured data about faculty research and scholarly activities for one institution, with the intention of repurposing the information in a variety of ways. . . . These systems give an overall picture of the research and scholarly enterprise of an institution"[9] and "RIM is used to refer to the integrated management of information about the research life-cycle, and about the entities which are party to it. . . . The aim is to synchronize data across parts of the university."[10] These descriptions all emphasize the role of the RIM system in aggregating data about research activities for an institution.

Although these definitions emphasize the role of aggregating *institutional* research outputs, other descriptions depart from this assumption. For example, Stvilia et al. define RIM systems as "types of information systems that manage and provide access to researchers' authored content and identity information and related services," including institutional systems as well as products like Google Scholar and ResearchGate.[11] However, these latter products do not provide a comprehensive aggregation of institutional outputs, and it is difficult or impossible to extract metadata for institutional use. In another example, several documents from the University of California Academic Senate emphasize the role of RIM systems in the assessment of research productivity, asserting that RIM systems "are used to aggregate data and generate metrics and statistics for universities and other institutions," and that they are "also called academic analytics." These statements discount the full range of business uses that RIM systems serve.[12]

In each of our interviews, we also asked participants about what they call their local system(s), and we heard little consensus in their responses. Many of our informants described their system by its local branding, such as "Activity Insight" at Penn State. We also heard descriptors like Expert Finder System (EFS) and Faculty Information System. Some informants even rejected the idea that their system constituted a type of RIM system at all, despite use of a vended RIM product to store and manage data about institutional research outputs; it was viewed as not being a RIM system because it wasn't used for reporting or assessment purposes.

TOWARD A UNIFIED DEFINITION OF RIM

The use of specialized terms describing a single usage scenario like "Research Networking System" or "Faculty Activity Reporting system" emphasize the differences rather than the significant similarities between systems. We assert that while being able to name and describe these use cases is important, it is urgent that US practitioners recognize RIM systems as a single, umbrella product category, with the potential to support numerous use cases with an array of different products. This shared understanding is a necessary step toward working across institutional silos and developing a collaborative, cross-functional community of practice in the United States.

It is urgent that US practitioners recognize RIM systems as a single, umbrella product category, with the potential to support numerous use cases.

We offer the following definition to describe what we believe constitutes RIM practice, in support of multiple use cases:

> **Research information management (RIM)** systems support the transparent aggregation, curation, and utilization of data about institutional research activities.

By *transparent*, we mean that, ideally, researchers should be able to examine the data and understand the sources from which it is collected, particularly because as systems that collect information about research outputs, the data may be used—or misused—for research assessment and metrics. Principle 4 of the *Leiden Manifesto for Research Metrics* asserts that data collection and analytical processes should be open, transparent, and simple.[13] Transparency is essential for trust by the researchers themselves; a lack of transparency can contribute to skepticism and fear.[14]

Additionally, the word *institutional* emphasizes the scope of these systems. It allows us to distinguish between efforts to collect, understand, and use information about campus research activities from external platforms like Google Scholar or ResearchGate, which were designed with researchers as the audience. While popular and well-used products by researchers, these systems do not aggregate institutional content to support any of the documented use cases in our study. They don't even serve as scalable metadata sources, nor do they offer a comprehensive view of institutional research activities. Although the inclusion of the word *institutional* suggests metadata aggregation for a single institution, we believe it can also embrace multi-institutional aggregation, such as seen with Florida ExpertNet in the United States[15] or a national system like CRIStin in Norway.[16]

We encourage all stakeholders in the US research community to consider adopting this comprehensive definition of RIM systems. Without recognizing that our similarities are greater than our differences, we will continue to work in silos, never realizing the potential to reduce effort, costs, and best support institutional goals.

RIM SYSTEMS SUPPORT MULTIPLE USE CASES

This study has documented six discrete use cases supported by RIM systems.

- **Faculty activity reporting** refers to processes that support annual academic progress reviews and/or promotion and tenure (P&T) processes for academic staff.
- **Public portals** feature profiles of individual researchers affiliated within an institution to expose and promote their research and that of the institution.
- **Metadata reuse** is the repurposing of data stored in the RIM system to populate department or faculty websites or internal lists beyond the original purpose of the RIM system.
- **Strategic reporting and decision support** refers to the use of RIM data to support reports, visualizations, and actionable recommendations. It is like metadata reuse, but it requires significantly more processing, data mining, and analysis.
- **Open access workflows** seek to streamline the processes for researchers to deposit open content into institutional repositories, adding convenience, improving metadata, and increasing discoverability and access to research.
- **Compliance monitoring** refers to the tracking and reporting required to ensure both individual and institutional compliance with external requirements, such as those that may be associated with awarded grants.

The public portal and faculty activity reporting use cases are dominant, occurring in all five institutions, while the compliance monitoring use case was found in limited ways at only one institution. This finding contrasts sharply with practices elsewhere, particularly in Australia, the United Kingdom, and Europe, where RIM systems play an essential role in compliance monitoring and reporting for external mandates relating to research assessment and open access.[17] External mandates are currently weaker in the United States than in these locales. Without these requirements driving practice, the US landscape has instead developed in a characteristically complex and decentralized manner, resulting in multiple systems and silos of practice. As a result, RIM practitioners supporting one business use may have little familiarity with other practices within their institution.

TABLE 1. Use cases represented in US case study narratives

Use cases	Texas A&M	Penn State	Virginia Tech	UCLA	Miami
Faculty activity reporting	Decentralized	✓	✓	✓	Decentralized
Public portal	✓	✓	Launching soon	✓	✓
Strategic reporting and decision support	✓	✓	✓	✓	✓
Metadata reuse	✓	✓	✓	✓	Planned
Open access workflow		✓	Planned	✓	
Compliance monitoring				✓	

Faculty activity reporting (FAR) practices are currently decentralized at two institutions in our study (Texas A&M and Miami), reflecting the historical decentralization of paper-based practices. However, the FAR use cases at UCLA, Virginia Tech, and Penn State reflect the institutional desire to centralize the management of activity reporting workflows, and we anticipate that more US institutions will seek to consolidate their FAR practices in the future.

All institutions in this study have an institutional repository (IR), and three (UCLA, Penn State, and Virginia Tech) are now using their RIM system to identify institutionally affiliated publications and support open access deposit into the IR. A fourth, the University of Miami, is implementing the Esploro product, which, when mature, is expected to operate as a hybrid repository/RIM system. These findings are congruent with previous research that has observed the merging of RIM and repository practices and products, a trend we expect to continue.[18]

DATA MUST BE FIT FOR PURPOSE

At its core, the data must be fit for purpose for each use case, which means it must be sufficiently complete, accurate, well-curated, deduplicated, granular, usable, machine-readable, transparent, and trustworthy to satisfy the requirements for that specific use.[19] Different business uses may have different data requirements; in all cases, institutions must assess if the data on hand is sufficient to meet the desired requirements. When data collated for one use is then extended to support another need, it may be found to be unfit for the new purpose. For example, data used for reporting must be as granular, deduplicated, complete, and transparent as possible. Regardless of use case, metadata librarians can play an important role in ensuring the data is fit for purpose.

RESEARCH INFORMATION MANAGEMENT IS AN ENTERPRISE EFFORT

In the course of documenting the key stakeholders and users for each RIM system in our case study narratives, we observed that the different use cases also engage different stakeholder communities—on and off campus. Table 2 provides a high-level summary of the stakeholders for each of the six use cases in this study, loosely organized by stakeholder relevance.

TABLE 2. Principal stakeholders represented in US case study narratives

Use case	Principal Stakeholders
Faculty activity reporting	• Faculty • Academic affairs units: Provost, deans, department heads • Academic human resources • The library (especially in Penn State example)
Public portal	• Faculty and other researchers • Campus units conducting research—including research centers and institutes • Office of Research • External audiences like government and business • Internal units that engage external audiences, such as advancement, corporate relations, and communications • The library
Strategic reporting and decision support	• Office of Research • Provost and other academic affairs units • Office of institutional research • The library • Faculty and other researchers
Metadata reuse	• Campus units • Faculty and other researchers • The library
Open access workflow	• The library • The faculty senate or administrative body responsible for local open access policies • Faculty and other researchers
Compliance monitoring	• Office of Research • Principal investigators

RIM systems are supported by a diversity of stakeholders from across the institution, which no doubt contributes to the multiple, siloed systems so indicative of the US RIM ecosystem. However, note that two groups appear as stakeholders across almost all use case categories: faculty/researchers and the library.

Faculty/researchers and their research activities are, after all, at the heart of the research information and institutional outputs collected. Institutional expectations that researchers will curate their own profiles universally end in disappointment, a reality that necessitates the engagement of library information professionals.

Library engagement with the FAR use case is particularly interesting. In three of the five case studies, the libraries had consciously distanced themselves from all FAR and evaluation activities due to concerns that library engagement in these activities could erode faculty trust in the library. In fact, in our interviews, two institutions explicitly described a "firewall" between library efforts and faculty review activities. However, this perspective was not universal. Two institutions in our study—Penn State and Virginia Tech—are actively involved in supporting FAR activities. Penn State is particularly notable because the Penn State Libraries-based Faculty Activity Management Services team is primarily responsible for the administration of the Activity Insight FAR system for the entire Penn State system, in close collaboration with academic affairs units. Virginia Tech is similarly engaged, but in closer partnership with other units. In both cases, the library plays a role not only in system administration, but also in data entry and management, allocating resources to support manual entry—which both reduces the burden on faculty and also helps ensure complete data. Interview participants from the libraries at both institutions expressed no concerns or negative consequences for the library in supporting these activities; instead, the library is seen as highly valued by faculty for its role in the process.

RIM Use Cases in the United States

In the process of documenting research information management practices at the five case study institutions, we identified six discrete use cases supported by RIM systems in the United States:

- Faculty activity reporting (FAR)
- Public portal
- Strategic reporting and decision support
- Metadata reuse
- Open access workflow
- Compliance monitoring

This section is intended to bring some order to the complex RIM ecosystem and provide a core vocabulary for discussion by categorizing and describing the functional goals and business needs that RIM systems may support. Our hope is that by exposing this broad array of use cases, practitioners and decision makers alike may better envision new ways their systems can support their institution's mission and goals.

Table 1 provides an overview of the use cases represented in the case studies, which are also described in depth in the institutional narratives.

FACULTY ACTIVITY REPORTING

This report uses the term "faculty activity reporting (FAR)" to describe use cases that support one or both of these workflows: annual academic progress reviews and promotion and tenure (P&T) processes for academic staff. While annual review and P&T are separate processes conducted at different intervals, they are similar in that they both collect much of the same information and require workflows to support creation, review by multiple parties, and decision-making. An electronic FAR system today utilizes an online user interface and the ability to produce the CV or reports required for review documentation, replacing former paper-based processes. If a FAR system is intended for both annual reporting and P&T, there will likely be some differences in the data collected:

- Annual reporting requires only activities for a recent reporting period—usually the last calendar year. It furthermore documents work in progress, such as submitted publications or grants.
- Promotion and tenure reports document the entirety of a scholar's accomplishments, requiring a comprehensive inventory of all relevant and completed scholarly activities over the course of the individual's career. Work in progress is rarely included.

Metadata implications

- FAR systems tend to collect only what is required for the individual faculty member to support the FAR process.
- As a result, the publications metadata may be free text, existing simply as a bibliographic citation cut and pasted from a CV.
- While most FAR systems do provide individual fields for basic citation metadata such as article title, journal name, list of authors, ISSN, and DOI, these systems typically *require* very few fields.
- It is common for FAR systems to include a separate publication record for each coauthor where there are multiple internal authors. This has implications for metadata reuse and reporting.

PUBLIC PORTAL

An online portal featuring profiles of individual researchers may be the most well-known and prevalent use case in the United States. These public portals, such as the examples documented at Penn State, Texas A&M, and UCLA, provide a way to expose and promote the research of individuals and the institution.

Although they may look identical at first glance, we believe it is important to make a distinction between two separate public portal sub-cases: research showcase and expertise discovery.

Both include researchers and their research outputs, but only the research showcase public portal aggregates research outputs, projects and grants, and activities to the unit level to create pages for departments, colleges, and centers and institutes, such as in Scholars@TAMU (Texas A&M) and the Penn State Research Portal.

Public portal—research showcase

A public portal is a research showcase if its intent is to show off the best, most influential work of the institution, both at an individual researcher level and at the level of units, projects, or facilities within the institution. By our definition, a research showcase highlights not only people, but also the structures in which they work: departments and colleges, research centers and institutes, and labs and project teams. This distinction enables users to identify a bibliography for a specific department and examine the strengths of a unit. This type of functionality can be particularly valuable to potential graduate school applicants or postdocs.

It is universally difficult to include the institutional hierarchy and researchers' multiple affiliations, as it is reliant on data pulled from other systems, such as HR or payroll, which may not document all units, job titles, or appointments. Institutional hierarchies are always unique. In addition, mapping out the institutional organization for a public portal requires identifying the relevant campus units and their hierarchical relationships within the institution as well as the mapping of individual researchers to the one or more units they are affiliated with. Depending on the situation, it may become important to include all of a researcher's publications and projects and affiliations/titles beyond what might be found in their HR record.

Metadata implications

- The metadata must include enough detail to aggregate to the levels necessary. In addition to each author, publication details typically include each internal author's home department, as well any other unit affiliations (project, institute, center, etc.) relevant for the publication.
- Since publications for a research showcase will be "rolled up" to a higher level in a hierarchy, publications must be deduplicated so there is only one record for each publication with multiple internal co-authors.

Public portal—expertise discovery

Expertise discovery use cases typically arise to support the need for researchers to find and connect with potential collaborators, particularly to catalyze and support interdisciplinary research, as seen in the UCLA Profiles example. Universities (or the state, as in the Florida ExpertNet example) may also want an expertise discovery portal so that people in industry can discover and connect with researchers to support local economic development.

Metadata implications

- Expertise discovery depends on the ability to effectively search the underlying metadata, both to support search engine optimization (SEO) for online searches and as the foundation for robust results and relevancy rankings within the portal itself. Therefore, it's important to provide as much text as possible about each expert's research.
- If the portal will feature only the people and their associated publications, grants, and so on, it isn't necessary to deduplicate publications—this becomes an issue as the institutional hierarchy is included, as discussed above with the research showcase.
- For text searches, complete publications metadata (including PIDs) may not be necessary; title and abstract for publications may be sufficient, with title, funder, and description being similarly adequate for grants. However, if the user wishes to filter results after an initial search, each potential filter field (publication year, department, etc.) must be a metadata element.
- The data in expertise discovery portals is often enriched with subject keywords and other metadata intended to help the user find what they're looking for. See the Florida ExpertNet sections of the University of Miami use case for an example.

It is exceptionally difficult to measure the success of expertise discovery efforts because users rarely notify the RIM system administrator that they identified collaborators by using the system.

METADATA REUSE

For our purposes, the metadata reuse use case generally means repurposing the data stored in the RIM system to populate department or faculty websites or internal lists beyond the original purpose of the system. Another common metadata reuse scenario occurs when multiple RIM systems exist, such as an expertise discovery portal and a FAR system. One system—hopefully, whichever is more complete—can supply data to the other as needed, avoiding duplicate data entry. This reuse is usually, but not always, facilitated by an API.

Metadata implications

- As long as the metadata includes a link between item and person, a faculty web page or a list of publications by that faculty member can be populated with just formatted text citations. However, more granular metadata allows for different citation styles, sorting, and filtering customized for each website or report.
- If data is passed between systems, attention must be paid to how the metadata structure in one maps to the other. Ideally, both systems use the same identifier (such as a DOI or ISBN) to minimize duplication.

STRATEGIC REPORTING AND DECISION SUPPORT

We consider strategic reporting and decision support as a separate use case from metadata reuse. Strategic reporting requires additional processing, data mining, and analysis to produce reports, visualizations, and actionable recommendations. Typical examples include:

- Network visualizations (for example, who collaborated with whom—both for people and for academic units)
- Geographic maps of collaborations
- Summaries of work by research area
- Analyses of changes in publication practices over time
- Reports of which internal authors are publishing in which journals (for library collection analysis)

For many institutions, it is when these additional business uses begin to emerge that it's discovered that the metadata in their RIM system, while sufficient for the original use case, is insufficiently granular to meet the requirements for the strategic reporting uses they have in mind.

Metadata implications

For strategic reporting, the data must be as granular, deduplicated, complete, and transparent as possible.

- Network visualizations for internal collaborations require author identifiers for each internal author on each publication, and internal affiliation for each author as part of their individual record or on each publication.
- Network visualizations for external collaborations require author identifiers for all authors as well as the affiliations for each author. For external authors, affiliations should be standardized and deduplicated.[20]
- Summaries of work by research area require a metadata element on each publication for each related research area.

- Any analysis done over time requires a publication year for each publication.
- Collection analysis reports require the journal ISSN.

OPEN ACCESS WORKFLOW

An open access (OA) policy is a policy in which institutions "require or request their researchers to provide open access to their peer-reviewed research article output by depositing it in an open access repository,"[21] a practice colloquially known as "green OA." Institutional OA policies are one of many ways that research institutions can take meaningful steps toward making scholarly content more openly and equitably available. However, institutions that have implemented or are considering OA policies must also consider a possible barrier: the authors themselves. Few researchers would argue with the principles of open access for their published work, but managing the details of getting the right version of each article into the right repository (and telling their institution about it) while managing license agreements and more adds another burden for each author.

The OA workflows described in the UCLA and Penn State case studies both seek to encourage researchers to follow the local policy while streamlining the compliance process. For example, the University of California has implemented a process where UC-affiliated publications metadata are harvested at scale, authors are emailed with a customized list of their publications that are not yet known to be openly available, and authors are offered a convenient way to deposit their publications into the institutional repository. This workflow eliminates the need for authors to enter publications metadata details themselves, as this information has already been collected on their behalf.

Metadata implications

- The OA workflow use case requires detailed metadata. In addition to author information that indicates author order and/or the corresponding author (if relevant to the policy), a DOI is vital for checking whether an item has already been deposited to the institutional repository or is openly available elsewhere.
- Other PIDs and metadata are also useful. Journal ISSN is used for checking publisher policies on green OA deposit. Publication year may be necessary for determining whether an item is subject to the OA policy. ORCID identifiers support author disambiguation.
- Publication records should be deduplicated to avoid situations where the same article is deposited by multiple authors, a waste of time for researchers and librarians alike.

COMPLIANCE MONITORING

In the United States, federal mandates are increasingly requiring open access to publicly supported research. Private funders like the Bill & Melinda Gates Foundation also have open access policies that require open access for work that their organization supports. In the course of this study, only the California Digital Library mentioned compliance monitoring as a use case, and only for certain entities, particularly the Lawrence Berkeley National Laboratory, which is an entity under the US Department of Energy but managed by the University of California. Compliance monitoring is already an important use case in other national environments like the United Kingdom, and we believe this usage scenario is likely to grow in importance in North America as well. We observe a growing interest in using the RIM system in two specific ways:

- Tracking which publications and datasets are associated with a grant
- Ensuring compliance with external open access requirements (which requires the association

of research outputs with the supporting award)

We encourage readers at institutions with an existing RIM system to share this potential use case with others at their institutions, particularly those with grant awards from NIH (which requires PubMed Central deposit)[22] and NSF (NSF-PAR deposit).[23] RIM systems populated with high-quality data can offer a time-saving alternative to laborious spreadsheets used today for monitoring and reporting on compliance with these mandates.

This use case may relate to open access, but it's not about facilitating deposit, as we detail in the OA workflow use case. Instead, this use is about making sense of the complex regulatory environment and the people, grants, and publications within it. It's also about tracking subcomponents of the university community: only some publications, resulting from specific grants, by specific principal investigators are necessary to monitor, and in relation to specific funder requirements. And while local OA policies at US institutions seldom include any consequences for noncompliance, failure to comply with a funder's mandate could result in a loss of funding.

Metadata implications

- Publications metadata must include the grant(s) that funded the research and an element to indicate that the item has been deposited in the required repository (or repositories, if funded by multiple institutions).

RIM System Framework

Each institution described in this report has its own combination of components that make up its RIM system(s). In order to consistently and comparatively describe the RIM systems documented in the case studies, as well as to provide readers with a scheme for understanding system components, we offer the following RIM System Framework in figure 1. This framework is divided into three high-level sections:

- Data sources
- Data processing (including storage)
- Data consumers

The RIM System Framework is intentionally shaped like an hourglass, representing the funnel of information into a core RIM system database and then out again in service of institutional business uses. The three discrete sections are color coded to illustrate their distinctions.

Data sources exist at the top of the funnel and refers to information flowing into the RIM system: data collected from outside the RIM system from both external and internal sources, such as publications and HR data. It is also often comprised of local knowledge that must be manually entered, such as research impact statements or research institutes that exist outside traditional academic unit hierarchies represented in HR systems.

Data processing constitutes the middle of the model, including not only the main RIM data store, but also the processes above it—the publication harvester, ETL processes, and metadata editor—that enable the transfer, cleaning, and enrichment of metadata into the RIM data store. At the bottom of the data processing section, below the data store, are the data transfer methods used to export the data in support of the various RIM use cases.

Data consumers at the bottom of the model refers to the outputs from the RIM system in the form of workflows serviced, metadata shared, or reports created, as detailed in the six use cases in this study.

We use this framework to document the functional and technical components of each RIM system in the five case studies detailed in Part 2. Note that in examples involving licensed or open source products, the framework documents features being used in the specific example rather than all the available features of the product. It is interesting to observe that the Virginia Tech implementation of Symplectic Elements uses the Elements product to support many of the system components, including metadata harvesting and as a data store. Texas A&M likewise uses Elements but only as a metadata harvester, reflecting different institutional decisions and the array of options available to RIM implementers. Not every system includes all components in the RIM System Framework, but each component is important to at least one of the RIM use cases identified.

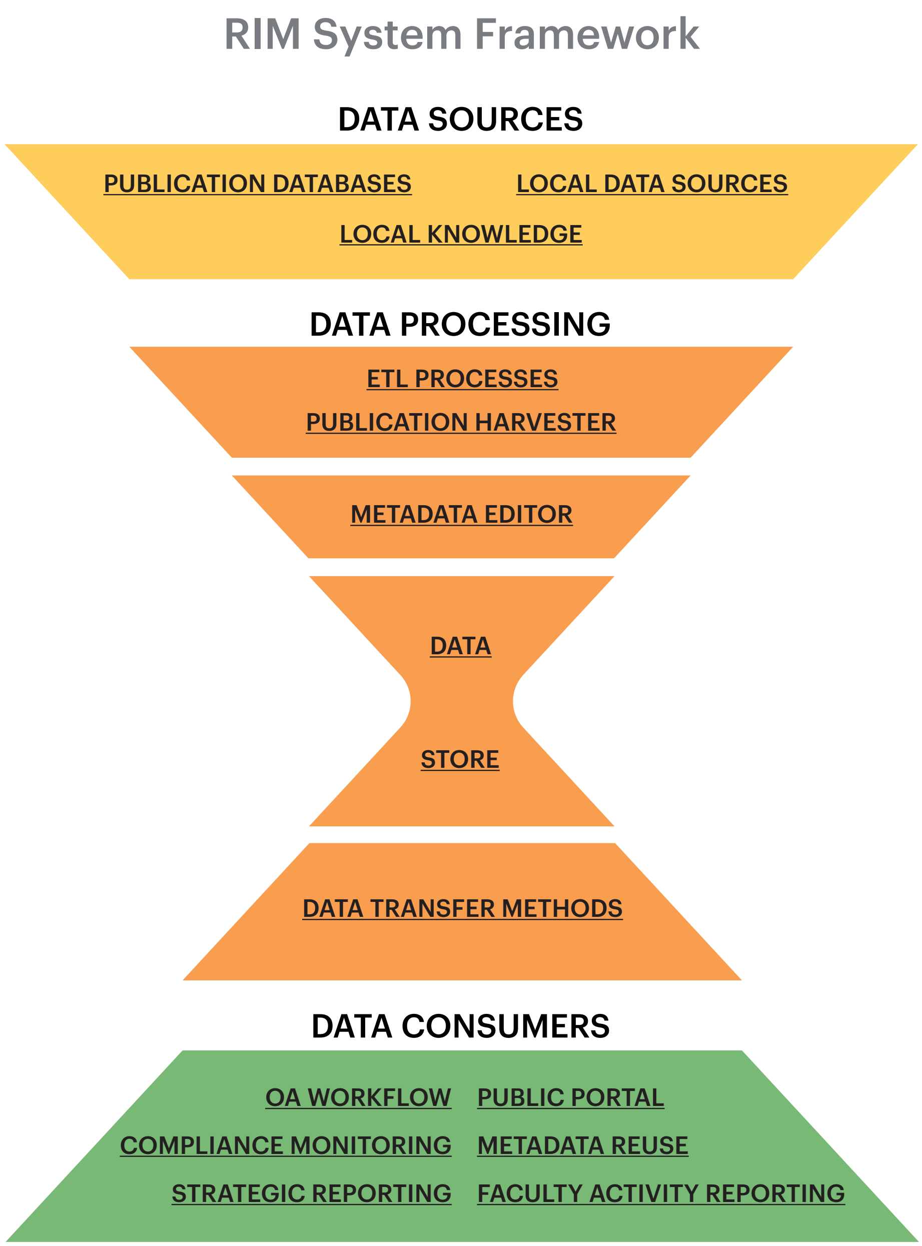

FIGURE 1. RIM System Framework

DATA SOURCES

Data sources, the top segment of the RIM System Framework, identify the information that must be aggregated from other domains. Depending upon the use case, this may include:

- Person name(s) and job title(s)
- Publications
- Patents
- Grants and projects
- Equipment
- Institutional units and their hierarchical relationships
- Instructional history
- Honors and awards
- Service activities
- Media reports
- Statements of impact

These metadata elements are then sourced from publication databases, local data sources, and local expertise and knowledge, as discussed below.

Publication databases and indexes

The vast majority of faculty members and researchers consider publications as their primary research output. Although some RIM systems—notably those used for faculty activity reporting—require the manual entry of publications information by individuals, most RIM systems rely on the indexes and databases licensed through the institution's library (e.g., Scopus or Web of Science) or freely available (e.g., PubMed Central) as the data source to populate individual profiles, with little effort required by researchers.

In each case study narrative, we list specific databases when they are the only or most prominent publication source for the RIM system. However, most systems rely on records from multiple databases.

Local data sources

Every academic institution is awash in data linked to people. Some of the most common examples include:

- Employees and their rank and job classification, contact information, affiliations, history with the organization, etc.
- Grants awarded by external funding agencies (as well as grants proposed but not funded), linked to the principal investigators
- Instructional history, including courses taught, linked to instructors

In each case study example in Part 2, we list the data category, such as Human Resources for data about employees, with additional details in the Metadata and sources section of the case study.

Local knowledge

Unfortunately, not all data one might want in a RIM system is readily available in another local system. Expected examples include data individuals must enter for themselves, like research or impact statements.

Note that organizational relationships and hierarchies often, perhaps surprisingly, fall into the local knowledge category. Our interviewees almost universally reported that data about the institutional hierarchies is elusive, incomplete, heterogeneous, and perpetually changing—and often completely unlinked to the people affiliated with these units. This is particularly true for nonacademic research centers and institutes. This presents a significant challenge for institutions seeking to support the "Public portal—research showcase" use case, where it is highly desirable to support reputation management for institutional units as well as their affiliates. Institutional hierarchies then become a form of local knowledge that must be pieced together by RIM administrators, and the RIM system itself may become the principal aggregator of this information, proving useful to users within and outside the organization.

We did not attempt to document all of the data derived from local knowledge for each use case, but you will find examples in the metadata and sources section of each case study.

> Our interviewees almost universally reported that data about the institutional hierarchies is elusive, incomplete, heterogeneous, and perpetually changing—and often completely unlinked to the people affiliated with these units.

DATA PROCESSING

The Data processing section of the RIM System Framework documents how that information is captured from other data sources, transformed, and stored for later use.

At its initial launch, the RIM system may be populated with information by exporting data from the identified sources, using queries and refinement tools to transform the data into a format that can be consumed by the RIM system, and importing each content type. However, in order to maintain current, representative information, these processes must be ongoing, and automated methods of adding and updating the system's data are required.

Publication Harvester

Metadata harvesting is the automated collection of metadata from one or more sources. A publication harvester allows the regular and automated updates of publications authored by researchers in a RIM system, drawing content from one or more publication databases and indexes, such as PubMed Central, Web of Science, Scopus, and many others. Some are freely available, and others are available as a source if the institution subscribes to the database. Harvesters attempt to identify and merge duplicate records from multiple sources, aggregating them into a common data structure and alerting users to potential duplicates that couldn't be automatically merged.

We documented four publication harvesters in the case studies:

- **Symplectic Elements** by Digital Science harvests from multiple databases and alerts the user to potential matches. For most databases, the user must review and claim suggested publications to add them to the RIM system. Elements can be configured to automatically import items from some databases with no claim required.
- **Elsevier Pure** also harvests from multiple databases and alerts the user to potential matches. If the client subscribes to Elsevier's Profile Refinement Service (PRS), Elsevier PRS curators match the institution's list of persons to the Scopus database and ensure that correct Scopus Author IDs are assigned and matched to the Pure record. Scopus records are then added automatically to Pure for the researchers covered by PRS.
- **The Esploro** product from Ex Libris uses a process called Smart Harvesting to automatically import publications metadata from Ex Libris's Central Discovery Index.
- **The open source Profiles RNS** system uses the Profiles RNS Author Disambiguation Engine to harvest from PubMed and works without the use of PIDs.

Publication harvesters benefit from network scale persistent identifiers (PIDs) for persons to support name disambiguation. (Profiles RNS is the exception to the rule.) ORCID is a vendor-agnostic identifier for researchers used by many publishers, funders, and institutions worldwide, which can also be linked to other proprietary identifiers such as the Scopus Author ID and ResearcherID. Institutions that rely on publication harvesters typically include as many person PIDs as possible with each researcher's RIM system profile in order to improve the quality of metadata harvesting.

More than one of the case study institutions considered developing their own harvester to pull publications from target databases but determined that the costs of maintaining a custom harvester as databases change outweighed the licensing costs.

For each case study, we identify the harvester used. For the few cases with no harvester, we specify the existing internal source used and/or indicate whether users or system administrators can import or manually add publications.

ETL processes

ETL stands for Extract, Transform, Load, and is a general term for computing processes that take data from one source, clean and crosswalk (mapping elements from one schema to another) as needed, and add or merge the data into a target database.[24] ETL processes cannot be accomplished with a vended RIM solution. Even where the product containing the source data is the same (Workday or PeopleSoft for HR data, for example) local business rules inevitably require ETL processes that are institutionally unique. It is important that an institution planning a RIM system also plan for the analyst and developer time needed to fit internal data to this new purpose.

For each case study in Part 2, we note the metadata elements that will be pulled from the internal sources for use in the RIM system in the Metadata and sources section of the case study.

Metadata editor

Although most data in a RIM system is aggregated from other sources, manual editing will always be necessary. The metadata editor is the interface that allows users to create, read, update, and delete information.

For example, the metadata editor may support functions such as:

- Claiming or disclaiming publications suggested by the publication harvester.
- Importing or manually adding publications and other types of research outputs that are not captured by the harvester.
- Adding and maintaining data available only as "local knowledge." For example, most institutions don't have a single list of all centers and institutes that can be easily imported.
- Tools for batch updates, such as the ETL processes for internal data.
- Tools for controlling access to the data store.

The metadata editor might exist as a very simple form enabling an individual or delegate to add biographical information, or it may encompass a much more complex range of activities for adding and editing the metadata in the data store. The metadata editor is typically part of either the data store or the harvester used by the RIM system.

For each case study, we identify the product for the metadata editor or note that it is a custom editor.

Data store

Every RIM system needs a place to store the data it aggregates. The data store might be part of a licensed product but may also be a bespoke database developed and maintained by the institution. If the latter, the institution must plan for a database administrator's time to set up and maintain the database.

For each case study, we identify the data store by either the product name or the name that the RIM team uses for it. For example, Texas A&M refers to their internally developed data store as simply the MySQL database. Florida ExpertNet's custom solution is a Microsoft SQL Server database that is part of the Knowledge Management System, or KMS.

Data transfer methods

In order to reuse the data stored in a RIM system, some method must exist for extracting it. The RIM systems we encountered typically included an application programming interface (API), allowing developers to write code that extracts data from the RIM system. Some RIM systems also allow data analysts to query the database directly with Structured Query Language (SQL), and some include reporting or exporting tools that are accessible to "power users." Any of these data transfer methods might be leveraged to meet the RIM use cases for the system. Note that for data used for strategic reporting and decision support uses, some ETL processing is expected.

For each case study, we identify the API and/or other methods used to extract data from the data store.

DATA CONSUMERS

Once the data has been collected and transformed, it can be extracted and used to support such uses as faculty activity reporting, OA workflows, metadata sharing, and report or dashboard creation. These uses are documented more fully in the previous RIM Use Cases section of this report, and examples are provided in the Part 2 case study narratives.[25]

Recommendations for Institutional Leaders

In the course of documenting the five RIM case studies, we have synthesized six high-level recommendations for US institutional leaders and decision makers.

INVEST IN INSTITUTIONAL DATA CURATION

Data must be fit for purpose for the use case it is serving, and it is essential to recognize that different business uses have different data requirements, as described above. Ideally, data collected for one use case can support many business needs, optimizing institutional investments and reducing duplication of effort. However, because of the differences of what constitutes sufficient data quality for different uses, efforts to extend RIM data to support additional needs can reveal data unfit for that new purpose, most especially for strategic reporting and decision support. RIM systems can provide significant insights into institutional activities, strengths, collaborations, and areas for additional investment and promotion—but only if the data is sufficiently granular, deduplicated, complete, and transparent.[26]

Quality data that is fit for purpose doesn't just happen—it requires significant institutional investment. It should not be outsourced to faculty who have few incentives to maintain thorough records of their activities, particularly in granular formats. Instead, we recommend that institutions rely upon librarians with metadata expertise to curate RIM system data, as observed at Penn State and Texas A&M. Additional manual metadata curation is particularly needed for humanities content, as it is underrepresented in publication data sources and is a pain point for research institutions worldwide.

Quality data that is fit for purpose doesn't just happen—it requires significant institutional investment.

Furthermore, faculty should always have access to data about their outputs and be able to review and ensure accuracy.

SUPPORT ADOPTION OF PERSISTENT IDENTIFIERS (PIDS)

A persistent identifier (PID) is a digital identifier that is globally unique, machine resolvable, and is used to reliably identify, verify, disambiguate, and locate a resource. Persistent identifiers are used to refer to a multitude of entities in the research ecosystem, such as publications, datasets, persons, organizations, projects, equipment, software, and much more.[27]

In an environment where digital information about research outputs and their contributors is distributed across the network, machine readable, interoperable PIDs such as DOIs and ORCID iDs enable the identification and harvesting of metadata, as well as supporting linkages between different types of entities. This facilitates an improved understanding of research relationships, offering the ability to link publications, to persons, to organizational affiliations, to grants, to funders, and much more. Widespread adoption of PIDs is crucial for metadata harvesting at scale.

PIDs also offer the promise of reducing the data entry burden on researchers. For instance, today researchers can create a biosketch for NIH and NSF grant proposals using the SciENcv biosketch tool that is integrated with ORCID, facilitating data transfer rather than rekeying.[28]

Libraries and publishers have long understood the value of PIDs and have been active proponents of their adoption. However, broader support by more stakeholders within research universities will bolster their adoption, which in turn will result in benefits to researchers as well as institutions. We encourage institutional leaders to inquire about the status of ORCID adoption at their institution and to encourage widespread adoption by researchers and integration into university systems, including, but not limited to, the RIM system. Institutions can also play a powerful role in following and supporting emerging PIDs like the Research Organization Registry (ROR), which can help with institutional name disambiguation for institutional rankings and benchmarking. There are many more PIDs emerging today that will help support a global system of scholarly information sharing, and more PIDs for more entities with more linkages will provide us all with better data.

DON'T EXPECT A TURN-KEY SYSTEM

RIM system vendors may emphasize that their system is easy to set up and can be operational within a matter of days. This promise can lead to disappointment when the implementation more typically takes weeks and months. So why does the implementation take so long?

The short answer: institutional data is rarely fit for purpose for reuse in a RIM system.

It can be difficult and time consuming to extract, transform, and load information from a local data source into the target RIM system.

For example, to implement any RIM system, the institution must start by identifying the people to be included in the system. Even if the decision seems simple, such as "all tenure-line faculty," this set of data may not be easily extracted from a local system(s), and it may also be highly complex, with multiple job titles and affiliations, zero percent appointments, and inconsistent abbreviations. As more researchers are added in different job categories, the complexity compounds. It can be difficult and time consuming to extract, transform, and load that information from a local data source into the target RIM system. Furthermore, implementation of the institutional RIM system can expose inconsistencies and gaps in existing campus data that will require remediation.

Organizational units and their hierarchical relationships also present a near-universal headache for RIM implementations because the institutional hierarchy may not be adequately captured in any other campus system, a situation that is particularly true for research institutes and centers. Information may need to be pulled from multiple systems and then harmonized. Abbreviations may need to be spelled out. And the relationship between researchers and their affiliated unit(s) may also be unclear.

And, of course, the process does not end with the initial go-live. The information will be out of date almost as soon as the system launches, so dedicated resources are required to maintain both the system and the data within it.

SUPPORT CROSS-FUNCTIONAL TEAMS

Research information management practices demonstrate growing operational convergence, as campus stakeholders must increasingly partner with others across the institution. This need is increasingly imperative as the skills required to successfully manage and use a RIM system are diverse, including expertise in metadata management, technical support, project management, communications and marketing, and research and reporting. There is also the need for significant time investments to ensure buy-in from multiple stakeholders.

In today's universities, we need our systems to be technically interoperable and our people to be socially interoperable. As defined in a previous OCLC Research report, social interoperability is the creation and maintenance of working relationships across individuals and organizational units that promote collaboration, communication, and mutual understanding.[29]

In today's universities, we need our systems to be technically interoperable and our people to be socially interoperable.

Virginia Tech is an outstanding exemplar of this approach. Throughout the case study, there are examples of listening, learning, responding to new challenges, and adjusting near-term objectives in service of institutional goals. There, the result of strong institutional collaboration and social interoperability is a single RIM system that supports multiple business needs and institutional objectives.

Without these close collaborations, multiple systems serving single use cases can emerge, often duplicating effort without reaping the full rewards of RIM implementation. Furthermore, inattention to social interoperability can make it difficult to gain access to institutional data (such as HR information) essential for RIM system management. Explicit executive sponsorship of RIM efforts—particularly when expressed in concert by multiple institutional leaders—can help facilitate cross-unit social interoperability.

INVEST IN DEDICATED PERSONNEL

As suggested in the sections above, RIM systems will not run themselves, and institutional investment beyond the purchase of a vended product is essential. We encourage institutions to realistically consider staffing needs in advance or else risk being disappointed in the results and usability of their investment. These are considerable investments. For example, two-thirds of the responding institutions to a global survey of RIM practices reported at least two full-time staff members supporting RIM activities.[30]

INCLUDE RESEARCH INFORMATION IN ENTERPRISE DATA GOVERNANCE EFFORTS

Institutions have long dedicated significant effort and resources in the meticulous management of student registration and academic history data. The collection and management of research information, particularly information about publications, datasets, and other forms of knowledge production, has, until recently, been managed at the individual level as content confined to the pages of a faculty CV.

Today, advancing technologies, standards, and networked information offer new ways to collect and manage this information at scale. A well-curated institutional bibliography, when combined with other internal and external data sources, can offer significant and granular insights into the relative strengths, growth, and opportunities for the institution. Publications metadata is a university asset, and like student, finance, and HR information, it should be included in campus-wide data governance and stewardship efforts. Ideally the metadata about research outputs should be collected and stored centrally, where access is facilitated and controlled, providing a convenient decision support platform for academic and administrative decision makers. This is exemplified in the Virginia Tech case study.

In addition, the negotiated license agreement for any product used should ensure that the institution retains its curated data if it discontinues use of that product.[31]

Publications metadata is a university asset, and like student, finance, and HR information, it should be included in campus-wide data governance and stewardship efforts.

CONCLUSION

This report summarizes the findings and recommendations synthesized from the examination of RIM practices at five US institutions. It introduces the RIM System Framework to facilitate the conceptualization and comparison of RIM system components, documenting the flow of data from external sources into the RIM data store and out again to support numerous uses. This report also describes six discrete RIM use cases identified in the course of the project. It offers a unified definition of research information management that embraces the diverse and siloed practices within US institutions and provides concise recommendations for institutional leaders who are seeking improved workflows, decision support tools, and clear return on investment.

The second component of this report series, *Research Information Management in the United States: Part 2—Case Studies* provides the evidence that supports the findings in this report through in-depth narratives about the RIM practices at five US research institutions. We encourage readers to also review Part 2 as it documents the history, use cases, and RIM system components used at each institution, as well as the roles of different stakeholders. Examination of these detailed narratives can help readers deepen their understanding of the challenges of RIM in the United States.

ACKNOWLEDGMENTS

The authors extend special thanks to our interview participants who generously shared their experiences and expertise with us for this investigation. This includes the 39 people interviewed for this study who also provided input on the draft case studies:

California Digital Library

- Katie Fortney, Copyright Policy and Education Officer
- Justin Gonder, Senior Product Manager, Publishing
- Lisa Schiff, Associate Director, Publishing, Archives, and Digitization
- Catherine Mitchell, Director, Publishing, Archives, and Digitization
- Alainna Wrigley, Publication Management System Coordinator

Florida ExpertNet

- Rebecca Augustyniak, Executive Director, University of Florida, Center for Information Management and Educational Services
- Amy Finley, Director, University of Florida, Center for Information Management and Educational Services

Penn State University

- Dan Coughlin, Head, Libraries Strategic Technologies and Interim Head of Research Informatics and Publishing, University Libraries
- Ana Enriquez, Scholarly Communications Outreach Librarian, University Libraries
- Nicole Gampe, IT Manager, Libraries Strategic Technologies, University Libraries
- Robyn Reed, Biomedical Informatics and Emerging Technologies Librarian, College of Medicine
- Michelle Hutnik, Director of Research Analytics and Communications, Office of the Senior Vice President for Research

Texas A&M University

- Jack Baldauf, Interim Vice President for Research, Office of the Vice President for Research
- David Carlson, Dean, University Libraries
- Doug Hahn, Director, Library Applications and Integration, University Libraries
- Bruce Herbert, Director, Office of Scholarly Communication, University Libraries
- Dong-Joon Lee, Research Information Systems Librarian, University Libraries
- Paula Sullenger, Associate Dean for Information Resources, University Libraries

University of California, Los Angeles (UCLA)

- Martin Brennan, Scholarly Communication Education Librarian, Library
- Jennifer Chan, Scholarly Communication Librarian, Library
- Anne Skinner, Administrative Director, UCLA Clinical and Translational Science Institute
- Heather Small, Project Lead, Opus, Academic Personnel Office

University of California, San Francisco (UCSF)

- Brian Turner, Product Director, Research Technology, UCSF Clinical and Translational Science Institute

University of Miami

- Leah Bamford, Senior Manager, Application Systems Development, Office of the Vice Provost for Research & Scholarship
- Angela Clark-Hughes, Director of the Rosenstiel School of Marine and Atmospheric Sciences Library, Libraries
- Charles Eckman, Dean and University Librarian, Libraries
- Elizabeth Gushee, Associate Dean, Digital Strategies, Libraries
- Ali Mosser, Senior Manager, Research Development and Education, Office of the Vice Provost for Research & Scholarship
- Eduardo Prieto, Integrated Systems Librarian, Libraries
- Elliot Williams, Digital Initiatives Metadata Librarian, Libraries

Virginia Tech

- Jack Finney, Vice Provost for Faculty Affairs, Office of the Executive Vice President and Provost
- Ben Greenawald, Data Scientist, Office of Analytics and Institutional Effectiveness
- Julie Griffin, Associate Dean, Research and Informatics, Libraries
- Abhay Joshi, Director for Academic Data Analysis and Visualization, Office of Analytics and Institutional Effectiveness
- Barbara Lockee, Faculty Fellow, Office of the Executive Vice President and Provost
- Gail McMillan, Director, Scholarly Communication, Libraries
- Andi Ogier, Director, Data Services, Libraries
- Ginny Pannabecker, Director, Research Collaboration and Engagement, Libraries
- Philip Young, Institutional Repository Manager, Libraries

We also extend our gratitude to external reviewers:

- Ruth Allee, retired from Northwestern University
- Meris Longmeier, The Ohio State University
- Mark Paris, Brandeis University

This report was also facilitated by a three-month research leave granted to Jan Fransen by the University of Minnesota Libraries. Thanks to University Libraries colleagues Dean Lisa German and Associate University Librarian John Butler (retired) for their encouragement and support. Special thanks to Caitlin Bakker and the entire Research Information Management Support Team who continued to move Minnesota's RIM system forward during Jan's leave.

Several OCLC colleagues provided guidance and support in the preparation of this report. Ixchel Faniel, Brian Lavoie, and Constance Malpas provided useful input on the project proposal. Kem Lang, OCLC Librarian, tirelessly helped us with our literature review. Annette Dortmund, Ixchel Faniel, Brian Lavoie, Constance Malpas, and Andrew Pace reviewed final manuscript drafts, providing invaluable external perspectives. Brian Lavoie and Chela Weber were likewise generous colleagues as we workshopped our ideas for the RIM System Framework.

The report could not have been published without the significant efforts of the OCLC Research publishing team, including Erica Melko, Jeanette McNicol, and JD Shipengrover.

Finally, our work was made possible by the senior leadership of OCLC; we wish to particularly thank Lorcan Dempsey, Vice President, Membership and Research, OCLC, for his continued support of this effort.

Author contributions:

- Bryant: Conceptualization, Data curation, Investigation, Methodology, Writing—original draft, Writing—review and editing, Project administration, Supervision
- Fransen: Investigation, Visualization, Writing—original draft, Writing—review and editing
- de Castro: Investigation, Writing—review and editing
- Helmstutler: Investigation, Writing—review and editing
- Scherer: Investigation, Writing—review and editing

NOTES

1. REF2021: Research Excellence Framework: https://www.ref.ac.uk/;

   Australian Government: Australian Research Council. 2021. "Excellence for Research in Australia (ERA)." Last modified 16 June 2021. https://www.arc.gov.au/excellence-research-australia;

   euroCRIS: https://www.eurocris.org/.

2. OCLC Research has previously published three reports on research information management practices:

   Bryant, Rebecca, Anna Clements, Carol Feltes, David Groenewegen, Simon Huggard, Holly Mercer, Roxanne Missingham, Maliaca Oxnam, Anne Rauh, and John Wright. 2017. *Research Information Management: Defining RIM and the Library's Role*. Dublin, OH: OCLC Research. https://doi.org/10.25333/C3NK88;

   Bryant, Rebecca, Annette Dortmund, and Constance Malpas. 2017. *Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information Management*. Dublin, OH: OCLC Research. https://doi.org/10.25333/C32K7M;

   Bryant, Rebecca, Anna Clements, Pablo de Castro, Joanne Cantrell, Annette Dortmund, Jan Fransen, Peggy Gallagher, and Michele Mennielli. 2018. *Practices and Patterns in Research Information Management: Findings from a Global Survey*. Dublin, OH: OCLC Research. https://doi.org/10.25333/BGFG-D241.

3. Bryant et al. *Practices and Patterns (see note 2)*.

4. Note that peer-reviewed publications represent just one of the many types of research outputs that today constitute the scholarly record. Other types of outputs include research data, code, data management plans (DMPs), and preprints.

5 See Bryant, Rebecca, Jan Fransen, Pablo de Castro. Brenna Helmstutler, and David Scherer. 2021. "Appendix: Semi-structured Interview Protocol." In *Research Information Management in the United States: Part 2—Case Studies, 46*. Dublin, OH: OCLC Research. https://doi.org/10.25333/qv1f-9e57.

6. The project team conducted interviews with individuals at the five case study institutions, plus individuals at Florida State University (which operates the state-wide ExpertNet system that Miami also participates in), the University of California, San Francisco (which supports UCLA Profiles), and the California Digital Library, which is part of the University of California system and supports the UC Publication Management System.

7. Azeroual, Otmane, and Joachim Schöpfel. 2019. "Quality Issues of CRIS Data: An Exploratory Investigation with Universities from Twelve Countries," 1. *Publications* 7(1): 1-18. https://doi.org/10.3390/publications7010014.

8. Schöpfel, Joachim, and Otmane Azeroual. 2021. "Current Research Information Systems and Institutional Repositories: From Data Ingestion to Convergence and Merger." In *Future Directions in Digital Information: Predicions, Practice, Participation—Chandos Digital Information Review*, 19–37. Edited by David Baker and Lucy Ellis. Chandos Publishing, Elsevier Ltd. https://doi.org/10.1016/b978-0-12-822144-0.00002-1.

9. Givens, Marlee. 2016. "Keeping Up With … Research Information Management Systems." *Keeping up with…* (Brief on trends in academic librarianship and higher education). American Library Association, 5 March 2016. https://www.ala.org/acrl/publications/keeping_up_with/rims.

10. Dempsey, Lorcan. 2014. "Research Information Management Systems: A New Service Category?" *Lorcan Dempsey's Weblog*, lorcandempsey.net, 26 October 2014. http://orweblog.oclc.org/research-information-management-systems-a-new-service-category/.

11. Stvilia, Besiki, Shuheng Wu, and Dong Joon Lee. 2019. "A Framework for Researcher Participation in Research Information Management Systems," 195. *Journal of Academic Librarianship* 45(3): 195–202. https://doi.org/10.1016/j.acalib.2019.02.014.

12. University of California. 2019. *University Committee on Academic Computing and Communications: Annual Report 2018-2019*. University of California. https://senate.universityofcalifornia.edu/_files/committees/ucacc/ar/ucacc-annual-report-18-19.pdf;

    University of California, Academic Senate. 2020. "University Committee on Academic Computing and Communications (UCACC), Thursday, October 15, 2020, Meeting Minutes" https://senate.universityofcalifornia.edu/_files/committees/ucacc/meetings/ucacc-10-15-20-minutes.pdf;

    Martone, Maryann E, Richard A. Schneider, Allegra Swift, Lisa Scheff, Jennifer Chan, and Catherine Mitchell. 2019. "Concerns Regarding the Use of Research Information Management Systems (RIMS)." *University of California Academic Senate* 12 March 2019. https://senate.universityofcalifornia.edu/_files/reports/rm-jn-mb-rims.pdf.

13. Diana Hicks, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. "The Leiden Manifesto for Research Metrics." *Nature* 520 (7548): 429-431. https://doi.org/10.1038/520429a.

14. For instance, contractual limits on information distribution and sharing have led to well documented concerns by US faculty as well as a statement from the American Association of University Professors. See:

    Basken, Paul. 2018. "UT-Austin Professors Join Campaign Against Faculty-Productivity Company." *Chronicle of Higher Education* 64 (21): 5. https://www.chronicle.com/article/ut-austin-professors-join-campaign-against-faculty-productivity-company;

    American Association of University Professors. 2016. "Statement on 'Academic Analytics' and Research Metrics." Washington, D.C. https://www.aaup.org/file/AcademicAnalytics_statement.pdf.

15 Florida ExpertNet (https://expertnet.org/) is described in detail in Part 2. See Bryant, Rebecca, Jan Fransen, Pablo de Castro. Brenna Helmstutler, and David Scherer. 2021. "Appendix: Semi-structured Interview Protocol." In *Research Information Management in the United States: Part 2—Case Studies, 46*. Dublin, OH: OCLC Research. https://doi.org/10.25333/qv1f-9e57.

16. To learn more about the Norway CRIStin infrastructure, see:

- CRISTIN: Current Research System in Norway. n.d. Accessed 30 Septemeber 2021. https://www.cristin.no/english/; and

- Wikiwand, Wikipedia reader. 2021. "CRIStin." Last updated 9 September 2021, at 23:36 (UTC). https://www.wikiwand.com/en/CRIStin.

17. Bryant, Rebecca, Anna Clements, Pablo de Castro, Joanne Cantrell, Annette Dortmund, Jan Fransen, Peggy Gallagher, and Michele Mennielli. 2018. *Practices and Patterns in Research Information Management: Findings from a Global Survey,* 40-45. Dublin, OH: OCLC Research. https://doi.org/10.25333/bgfg-d241.

18. This assertion is supported, for example, by these publications:

Schöpfel, Joachim, and Otmane Azeroual. 2021. "Current Research Information Systems and Institutional Repositories: From Data Ingestion to Convergence and Merger, 19." In *Future Directions in Digital Information*. Edited by David Baker and Lucy Ellis. Chandos Publishing, Elsevier Ltd. https://doi.org/10.1016/b978-0-12-822144-0.00002-1.

Castro, Pablo De, Kathleen Shearer, and Friedrich Summann. 2014. "The Gradual Merging of Repository and CRIS Solutions to Meet Institutional Research Information Management Requirements." *Procedia Computer Science* 33: 39–46. https://doi.org/10.1016/j.procs.2014.06.007;

Bryant, et al. *Practices and Patterns* (see note 17).

19. See also Azeroual, Otmane, and Joachim Schöpfel. 2019. "Quality Issues of CRIS Data: An Exploratory Investigation with Universities from Twelve Countries," 2. *Publications* 7(1): 14. https://doi.org/10.3390/publications7010014.

20. That is, if one author's affiliation says "Case Western" and another's says "Case Western Reserve University," the metadata should either be standardized to use one or the other, or both should point to a single, preferred organization record for the institution. The increasing and widespread adoption of the Research Organization Registry (ROR) in scholarly communications will also improve organizational disambiguation over time. Geographic map requirements are the same as network visualizations for external collaborations, and records for external organizations must include geographic location information. ROR includes location as part of an organization's identity;

For more information on ROR, see https://ror.org/.

21. Registry of Open Access Repository Mandates and Policies (ROARMAP), http://roarmap.eprints.org/.

22. The NIH Public Access Policy requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funding to PubMed Central immediately upon acceptance for publication. US Department of Health and Human Services. “NIH Public Access Policy.” https://publicaccess.nih.gov/.

23. The NSF also has a public access policy, effective 25 January 2016. See National Science Foundation: Where Discoveries Begin. n.d. “Public Access To Results of NSF-Funded Research.” Accessed 24 September 2021. https://www.nsf.gov/news/special_reports/public_access/index.jsp;

    It offers NSF awardees the opportunity to deposit their work in the NSF Public Access Repository (NSF-PAR,). See Nation Science Foundation: Where Discoveries Begin. n.d. “About the NSF Public Access Repository (NSF-PAR).” In Public Access to Results of NSF-Funded Research. Accessed 24 September 2021. https://www.nsf.gov/news/special_reports/public_access/about_repository.jsp.

24. Wikipedia provides a sufficient description at:

    Wikipedia. 2021. “Extract, Transform, Load.” Last modified 7 July 2021, at 15:03 (UTC). https://en.wikipedia.org/wiki/Extract,_transform,_load.

25 Bryant, Rebecca, Jan Fransen, Pablo de Castro. Brenna Helmstutler, and David Scherer. 2021. “Appendix: Semi-structured Interview Protocol.” In *Research Information Management in the United States: Part 2—Case Studies, 46*. Dublin, OH: OCLC Research. https://doi.org/10.25333/qv1f-9e57.

26. Azeroual et al. provides an excellent overview of the issue of data quality in RIM systems:

    Azeroual, Otmane, Gunter Saake, Mohammad Abuosba, and Joachim Schöpfel. 2020. “Data Quality as a Critical Success Factor for User Acceptance of Research Information Systems.” *Data* 5 (2). https://doi.org/10.3390/data5020035.

27. The PID Forum and OpenAIRE provide useful definitions of PIDs:

    - the PID Forum. 2021. “Persitent Identifier (PID) Definition.” Last updated 16 July 2021. https://www.pidforum.org/t/persistent-identifier-pid-definition/1502;
    - OpenAIRE. 2017. “What is a Persistent Identifier?” https://www.openaire.eu/what-is-a-persistent-identifier;

    The scholarly communications community also supports the annual PIDapalooza conference to share about advancements in persistent identifiers:

    - Pidapalooza. 2021. “Home.” https://www.pidapalooza.org/.

28. LYRASIS. 2020. “SciENcv and ORCID to Streamline NIH and NSF Grant Applications.” *LYRASIS Now* (blog), 18 April 2020. https://lyrasisnow.org/sciencv-and-orcid-to-streamline-nih-and-nsf-grant-applications/.

29. Bryant, Rebecca, Annette Dortmund, and Brian Lavoie. 2020. *Social Interoperability in Research Support: Cross-Campus Partnerships and the University Research Enterprise*, 2. Dublin, OH: OCLC Research. https://doi.org/10.25333/wyrd-n586.

30. Bryant, Rebecca, Anna Clements, Pablo de Castro, Joanne Cantrell, Annette Dortmund, Jan Fransen, Peggy Gallagher, and Michele Mennielli. 2018. *Practices and Patterns in Research Information Management: Findings from a Global Survey,* 67. Dublin, OH: OCLC Research. https://doi.org/10.25333/BGFG-D241.

31. We heard assertions from interview participants as well as others in our networks that proprietary products like Pure will not allow institutions to retain their own data if the license is discontinued. This is inaccurate and is instead a matter of institutional license negotiation.

For more information about our work related to research information management, please visit: **oc.lc/us-rim-project**

OCLC

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: 978-1-55653-224-5
DOI: 10.25333/8hgy-s428
RM-PR-216916-WWAE 2110