Research Information Management Practices at US Institutions

Rebecca Bryant, PhD
Senior Program Officer,
OCLC Research Library Partnership
bryantr@oclc.org  @RebeccaBryant18
https://orcid.org/0000-0002-2753-3881

Jan Fransen
Service Lead for Research Information Management Systems, University of Minnesota
fransen@umn.edu  @umjanlib
https://orcid.org/0000-0002-0302-2761
We acknowledge and celebrate the Indigenous people on whose traditional land and airways we meet, and pay our respect to the elders past and present.

If you are unsure whose land you are currently residing upon, we encourage you to visit native-land.ca.
As of 30 June 2020

A global network of libraries

- **Americas**: 9,033 members in 26 countries
- **EMEA**: 5,217 members in 61 countries
- **Asia Pacific**: 1,387 members in 20 countries
Devoted to challenges facing libraries and archives since 1978
- Aims to scale and accelerate library learning, innovation, and collaboration
- oc.lc/research

The OCLC Research Library Partnership supports collaboration and learning among a transnational group of research libraries
- oc.lc/joinRLP
OCLC Research publications on Research Information Management

Research Information Management: Defining RIM and the Library’s Role
Rebecca Bryant
DLC Research (US)
Anna Clements
University of St. Andrews (UK)
Carol Fiske
Rockefeller University (US)
David Groenewegen
Vrije Universiteit (Netherlands)
Elana Haggard
La Trobe University (Australia)
Holly Marcus
University of Tennessee-Knoxville (US)
Ruxandra Milasihour
Australian National University (Australia)
Melissa O’Keefe
University of Arizona (US)
Anna Paish
University of Arizona (US)
Jen Wright
University of Delaware (US)

Convenience and Compliance:
Case Studies on Persistent Identifiers in European Research Information Management
Rebecca Bryant, Anneta Dortmund, and Constance Malpas

Practices and Patterns in Research Information Management
Findings from a Global Survey
Rebecca Bryant, Anna Clements, Pablo de Castro, Joanne Centrell, Anneta Dortmund, Jan Fransen, Peggy Gallagher, Michele Masselli
Research Information Management in the United States

Project Team

Rebecca Bryant, PhD, OCLC Research co-PI
Jan Fransen, University of Minnesota co-PI
Pablo de Castro, euroCRIS & Strathclyde University
Brenna Helmstutler, Syracuse University
David Scherer, Carnegie Mellon University

oc.lc/us-rim-project
Project overview

Scope

5 US research universities with a diversity of:
- RIM use cases
- Products
- Scale & scope
- Stakeholders

Methodology

- Case study approach
- Semi-structured interviews with 39 individuals at 7 institutions

Outputs

- Forthcoming OCLC Research report, expected ~September 2021
- Webinars & conference presentations
- Blogs & discussions
Project goals

• Examine & document the breadth and depth of RIM practices at case study institutions in the United States
• Explore the important factors driving RIM adoption and use at US institutions
• Document
  o RIM use cases
  o A RIM framework of technical/functional components
  o Recommendations to institutions

oc.loc/us-rim-project
<table>
<thead>
<tr>
<th>Institutions</th>
<th>Use case(s)</th>
<th>Relevant products used</th>
<th>Are there implementations above the institutional level?</th>
</tr>
</thead>
</table>
| Penn State                                                                 | • Public profiles  
• Reporting  
• Faculty Activity Reporting  
• OA policy support | • Elsevier Pure,  
• Digital Measures FAR,  
• Home grown Researcher Metadata Database | All systems are implemented at the university system level, across 24 campuses |
| Texas A&M                                                                  | • Public profiles  
• Reporting  
• repository integration | • VIVO  
• Symplectic Elements  
• MySQL | no |
| Virginia Tech                                                              | • Faculty Activity Reporting,  
• Public profiles  
• repository integration  
• Reporting | • Symplectic Elements  
• Symplectic Elements Discovery module | no |
| UCLA (and University of California)                                        | • FAR (Opus)  
• public profiles  
• OA policy support/repository | • Home grown FAR system  
• Interfolio  
• Symplectic Elements  
• Profiles RNS | OA policy and publication management system at the system level |
| University of Miami                                                        | • Public profiles/repository  
• Public profiles  
• FAR | • Ex Libris Esploro  
• Elsevier Pure | Florida ExpertNet |

[oc.lc/us-rim-project]
The transparent aggregation, curation, and utilization of (meta)data about institutional research activities

RIM =

These systems might be called:

- Research Information Management System
- Researcher Profile System
- Expert Finder System
- Research Information System (RIS)
- Research Networking System (RNS)
- Faculty Activity Reporting (FAR) system
- Faculty Information System (FIS)
- Current Research Information System (CRIS) (if used, not used in the same way as in Europe)
Why is there so little agreement on what we call these things?

Disparate use cases
- Faculty activity reporting
- Public portal: Expertise discovery &/or Research showcase
- Supporting campus OA policies
- Strategic reporting & decision support
- Metadata reuse

Lots of stakeholders
- ... in a complex adaptive university system

Few national mandates
- RIM systems rarely used for compliance monitoring (at this time)
RIM

- Faculty Activity Reporting
- HR/Academic Affairs sectors to support workflows
- Home grown
- Open source, seeded by NIH
- New entrant, with institutional repository at the core
- CRIS systems
- Publishing sector
- Library sector
USE CASES
## Six Primary US RIM Use Cases

<table>
<thead>
<tr>
<th>Faculty Activity Reporting (FAR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Portal: Expertise Discovery &amp;/or Research Showcase</td>
</tr>
<tr>
<td>Open Access Workflow</td>
</tr>
<tr>
<td>Strategic Reporting &amp; Decision Support</td>
</tr>
<tr>
<td>Metadata Reuse</td>
</tr>
<tr>
<td>Compliance Monitoring</td>
</tr>
</tbody>
</table>
## RIM Use Cases at Case Study Institutions

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Texas A&amp;M</th>
<th>Penn State</th>
<th>Virginia Tech</th>
<th>UCLA</th>
<th>Miami</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty Activity Reporting (FAR)</td>
<td>decentralized</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>decentralized</td>
</tr>
<tr>
<td>Public Portal: Expertise Discovery &amp;/or Research Showcase</td>
<td>✔</td>
<td>✔</td>
<td>in progress</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Open Access Workflow</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Strategic Reporting &amp; Decision Support</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Metadata Reuse</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Compliance Monitoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>
What happens when you have multiple drivers, use cases, & stakeholders?

... And in a highly autonomous & decentralized ecosystem
“We have six or seven research profiling systems [on our campus]. That is duplication of service, for sure.”
In the US:
- FAR, public portals, and other systems/workflows are still largely separate
- While we documented institutions with centralized FAR systems, many units still maintain their own processes

Why does this matter?
- The data often can’t be aggregated for a consistent, enterprise-level view
- And data that is ”good enough” for profiles may be incomplete for other purposes (hello! reporting?)
- Faculty are still being asked to enter information into multiple systems.
RIM SYSTEM FRAMEWORK
Data Extraction

Local knowledge

Publication Databases

Publication Harvester

Publication Sources

ETL Processes

Metadata Editor

Data Store

Data Transfer Methods

Public Portal

Data Extraction

FAR workflow

Institutional Repository

Data Sources

Data Processing

Data Consumers
Publication Databases
Publication Harvester
Metadata Editor
Data
Store
Data Transfer Methods
Public Portal
Data Extraction
FAR workflow
Institutional Repository

Local data sources
ETL Processes
Data
Publication Databases
Local knowledge
Data Sources
Data Processing
Data Consumers
CASE STUDIES
PENN STATE UNIVERSITY
About Penn State University

• The Pennsylvania State University ("Penn State") is a state-related research-intensive university

• Enrollment:
  – ~75,000 undergraduates
  – > 15,000 graduate and professional students

• 24 campuses
<table>
<thead>
<tr>
<th>System</th>
<th>Key Features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elsevier Pure</strong></td>
<td>• Public portal: expertise discovery &amp; research showcase</td>
</tr>
<tr>
<td></td>
<td>• Public profiles for about 5,500 researchers</td>
</tr>
<tr>
<td></td>
<td>• Managed by research office &amp; medical college</td>
</tr>
<tr>
<td><strong>Activity Insight</strong></td>
<td>• Supports Faculty Activity Reporting (FAR) for &gt;7,000 faculty on 24 campuses</td>
</tr>
<tr>
<td></td>
<td>• Managed by library</td>
</tr>
<tr>
<td></td>
<td>• Library supports data entry</td>
</tr>
<tr>
<td><strong>Researcher Metadata Database (RMD)</strong></td>
<td>• Supports institutional open access policy/workshop</td>
</tr>
<tr>
<td></td>
<td>• Aggregates metadata from Pure &amp; AI</td>
</tr>
<tr>
<td></td>
<td>• Supports reporting &amp; data reuse</td>
</tr>
<tr>
<td></td>
<td>• Managed by library</td>
</tr>
<tr>
<td>RIM Use Cases Penn State</td>
<td>Penn State</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>Faculty Activity Reporting (FAR)</td>
<td>✔</td>
</tr>
<tr>
<td>• Library manages <a href="#">Activity Insight FAR system</a> for 24 campuses</td>
<td></td>
</tr>
<tr>
<td>Public Portal: Expertise Discovery &amp;/or Research Showcase</td>
<td>✔</td>
</tr>
<tr>
<td>• <a href="#">Elsevier Pure portal</a> includes 5,500 faculty profiles for all campuses and colleges</td>
<td>• Managed jointly by Office of Research and College of Medicine</td>
</tr>
<tr>
<td>Open Access Workflow</td>
<td>✔</td>
</tr>
<tr>
<td>• <a href="#">Home grown Research Metadata Database (RMD)</a> supports institutional OA policy</td>
<td></td>
</tr>
<tr>
<td>Strategic Reporting &amp; Decision Support</td>
<td>✔</td>
</tr>
<tr>
<td>• Research Metadata Database (RMD) supports reporting</td>
<td></td>
</tr>
<tr>
<td>Metadata Reuse</td>
<td>✔</td>
</tr>
<tr>
<td>• Research Metadata Database (RMD) supports data sharing</td>
<td></td>
</tr>
<tr>
<td>Compliance Monitoring</td>
<td></td>
</tr>
</tbody>
</table>
Simplified timeline for Penn State

- Initial rollout of Digital Measures Activity Insight (2009)
- SciVal licensed by OSVPR (2015)
- Pure added to OSVPR license (2016)
- Pure portal goes public (2018)
- Provost and Dean of Libraries charge Open Access task Force (2017)
- University continues to manage Activity Insight
- College of Medicine launches RNS Profiles (2011)
- Penn State Open Access Policy goes into effect (2020)
- Researcher Metadata Database launches (2021)
Penn State Activity Insight

- Scopus, Web of Science, PubMed, CrossRef
- Local knowledge
- HR, sponsored projects, registrar data
- Persons, grants, courses taught
- Activity Insight + CV Importer Tool
- Activity Insight API
- Public Portal
- Researcher Metadata Database
- Activity Insight
- Institutional Repository

Data Sources

Data Processing

Data Consumers
Penn State Pure

Scopus (automatic) + other publication databases

Profile Refinement Service + Pure

Local knowledge

HR data (central and medical school)

Pure Master List (Persons + Orgs)

Pure

Pure Web Services, exports

Pure

Researcher Metadata Database, SciVal

FAR workflow

Institutional Repository

Data Sources

Data Processing

Data Consumers
About UCLA

• University of California, Los Angeles (UCLA)
  – Public research institution
  – Enrollment: 31,000 undergraduate students & 13,000 postgraduate students in > 300 degree programs
  – One of 10 campuses of the University of California System

• University of California (system)
  – Largest higher education system in the US
  – Comprised by 10 research universities, including UCLA
  – In total, >250,000 students, 22,000 faculty members, 150,000 staff

• The California Digital Library (CDL) serves as a central library hub to support UC-wide efforts
## UCLA—3 distinct systems

<table>
<thead>
<tr>
<th>Opus:</th>
<th>UC Publication Management System (PMS)</th>
<th>UCLA Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Home grown system aggregating faculty information, coupled with</td>
<td>• Supports institutional OA policy for 22,000 faculty and thousands of other researchers</td>
<td>• Profiles RNS</td>
</tr>
<tr>
<td>• Interfolio systems for tenure/promotion &amp; annual reviews</td>
<td>• Elements for metadata harvesting</td>
<td>• 1,000+ public profiles of biomedical researchers</td>
</tr>
<tr>
<td>• Managed by UCLA personnel office</td>
<td>• Managed by California Digital Library</td>
<td>• Managed by Clinical &amp; Translational Science Institutes at UCLA and UCSF</td>
</tr>
<tr>
<td><img src="image" alt="Home grown" /></td>
<td><img src="image" alt="Interfolio" /></td>
<td><img src="image" alt="UCLA Profiles" /></td>
</tr>
<tr>
<td>Use Case</td>
<td>UCLA</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>-----------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Faculty Activity Reporting (FAR)</td>
<td>✓ Opus &amp; Interfolio systems on UCLA campus</td>
<td></td>
</tr>
<tr>
<td>Public Portal: Expertise Discovery &amp;/or Research Showcase</td>
<td>✓ UCLA Profiles includes only biomedical researchers. Supported by the UCSF campus</td>
<td></td>
</tr>
<tr>
<td>Open Access Workflow</td>
<td>✓ UC Publication Management System notifies researchers of publications eligible for OA deposit</td>
<td></td>
</tr>
<tr>
<td>Strategic Reporting &amp; Decision Support</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Metadata Reuse</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Compliance Monitoring</td>
<td>✓ Discrete uses to manage compliance with federal grant requirements, thru UC PMS.</td>
<td></td>
</tr>
</tbody>
</table>
Simplified timeline for UCLA

UC launches OA policy for faculty (2013)

CDL begins licensing Symplectic Elements (Dec 2013)

OA policy expanded to include all UC researchers (2015)

UC PMS roll out begins across all UC campuses, including UCLA and still ongoing today (2021)


Opus development begins (2013)

Opus database developed (2016)

UCLA licenses Interfolio RPT (2016)

Interfolio roll out, also expanding to include annual reviews

Entire campus now using Interfolio RPT (2020)

Expansion of biomedical profiles to other UC campuses (2017-)

UCSF Profiles launched (2010)

UCLA Profiles launched (2020)
UC Publication Management

Enables you to

- Manage your publications.
- Participate in the UC Open Access Policy.
- Increase the impact of your work.

Select your campus to get started:

- UC BERKELEY
- UC RIVERSIDE
- UC DAVIS
- UC SANTA BARBARA
- UC IRVINE
- UC SANTA CRUZ
- UCLA
- UC SAN DIEGO
- UC MERced
- UCSF
- LBL
- NON-UC

Managing your publications

- We'll scan the web for publications you've authored.
- Log in (at left) to review what we've found.
- Claim publications that are yours; reject those that aren't.
- Upload your manuscript for public display on eScholarship.

Resources and support

- Learn more about the UC Open Access Policy.
- Get answers to Frequently Asked Questions.
- Find out who to contact for additional support.

Logout notice

To protect your accounts from unauthorized access, please lock your workstation or exit your browser after logging out of this site.
TEXAS A&M
About Texas A&M University

• Public research university in College Station, Texas
• Enrollment:
  – > 56,000 undergraduate students
  – ~16,000 postgraduate & professional students
TAMU—Multi-component system

Scholars@TAMU

- Multiple components:
  - Elements: metadata harvester
  - MySQL: data store
  - VIVO (open source): public portal
  - Managed by library

Note:
- TAMU also conducts faculty reviews, but these are in several systems across multiple colleges
- We chose not to document decentralized FAR systems in this study
- However, do note that this decentralization seems common in the US at this time
### RIM Use Cases at Texas A&M

<table>
<thead>
<tr>
<th><strong>Faculty Activity Reporting</strong></th>
<th>• Currently no campus-wide system in place, but moving to Interfolio RPT.</th>
</tr>
</thead>
</table>
| **Public Portal: Expertise Discovery & Research Showcase** | ✔ | • Catalyze interdisciplinary collaboration  
• Showcase research to maximize the reputation of both the university and its faculty  
• Display outputs/people by unit (including centers) |
| **Open Access Workflow** | | |
| **Strategic Reporting & Decision Support** | ✔ | • Developed an authoritative data store for frequently requested information about faculty contributions to the University and assist leaders in making strategic decisions.  
• **Decision support dashboard for VP Research in development** |
| **Metadata Reuse** | in progress | • Campus is implementing Interfolio FAR; plan to populate FAR system with Scholars information |
| **Compliance Monitoring** | | |
Simplified timeline for Texas A&M

2013: Library begins exploring VIVO (2013)
2015: Tightening goals (2015)
2016: Licensed Elements and relaunched VIVO portal (2016)
2017: Custom portal (TAMU Scholars) launches (2017)
2019: Code for TAMU Scholars and from Duke Scholars make up VIVO Scholars (2019-20)
2020: Dashboard tool in development (2021)
Texas A&M

**Publication Databases**

- Local knowledge

**Data Sources**

- HR, sponsored projects, academic history data
- Persons, grants, courses taught, committee service

**Data Processing**

- MySQL
- Scholars Middleware, SQL queries
- Custom Editor

**Data Consumers**

- VIVO
- DSpace
- DSpace
- Data Extraction
- FAR workflow

**Publication Databases**

- Elements

**Data Sources**

- VIVO
- DSpace
- Data Extraction
- FAR workflow
CONCLUSIONS
### Metadata quality matters

- Must be complete, transparent if used for reporting
- Don't rely on faculty
- Humanities metadata gap

### Persistent Identifiers (PID) are essential infrastructure

- DOIs, ORCID, ROR, RAID, etc.
- Allow us to link:
  - Researcher
  - Publication
  - Grant
  - Project
  - Institutional affiliation
  - Collaborators
- Support a future of "enter once, reuse often"

### These are not turn-key systems

- EVERYONE struggles to manage the institutional hierarchy
- Inclusion of centers & institutes isn't easy

### Team-based, enterprise approach is best

- Array of skills—and social interoperability—needed:
  - Technical
  - Metadata
  - Project management
  - Communications
- Don't scrimp on dedicated staff
- Plan for multiple use cases to > buy-in and < faculty burden

---

**Recommendations**
We’d like to hear from you

1. What’s your response to our description of the US RIM ecosystem?
   • Is there anything that surprises you?
2. What seems most different (or similar!) from the European CRIS/RIM landscape?
3. Do you have any comments on the RIM System Framework?
4. Advice
   • What advice would you give to institutional leaders in the US?
   • To RIM practitioners?
References


