



AUDIT AND ADVISORY SERVICES

Research Data Management Audit Project No. 16-672

June 24, 2016

Prepared by:

Chad Edwards
Auditor-in-Charge

Reviewed by:

Approved by:

Jaime Jue
Associate Director

Wanda Lynn Riley
Chief Audit Executive

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

AUDIT AND ADVISORY SERVICES
Tel: (510) 642-8292

611 UNIVERSITY HALL #1170
BERKELEY, CALIFORNIA 94720-1170

June 24, 2016

Larry Conrad
Associate Vice Chancellor for Information Technology and Chief Information Officer
Office of the Chief Information Officer

Associate Vice Chancellor Conrad:

We have completed our audit of research data management as per our annual service plan in accordance with the Institute of Internal Auditors' Standards for the Professional Practice of Internal Auditing and the University of California Internal Audit Charter.

Our observations with management action plans are expounded upon in the accompanying report. Please destroy all copies of draft reports and related documents. Thank you to the staff of Research Information Technologies and the Library for their cooperative efforts throughout the audit process. Please do not hesitate to call on Audit and Advisory Services if we can be of further assistance in this or other matters.

Respectfully reported,

Wanda Lynn Riley
Chief Audit Executive

cc: Interim Executive Vice Chancellor & Provost Carol Christ
Vice Chancellor Paul Alivasatos
Vice Chancellor Scott Bidy
University Librarian Jeff Mackie-Mason
Director David Greenbaum
Manager Christopher Hoffman
Associate University Librarian Erik Mitchell
Senior Vice President and Chief Compliance and Audit Officer Sheryl Vacca
Associate Chancellor Khira Grisavage
Assistant Vice Chancellor and Controller Delphine Regalia

**University of California, Berkeley
Audit and Advisory Services
Research Data Management**

Table of Contents

OVERVIEW	2
Executive Summary	2
Source and Purpose of the Audit	3
Scope of the Audit	3
Background Information.....	4
Summary Conclusion.....	4
SUMMARY OF OBSERVATIONS & MANAGEMENT RESPONSE AND ACTION PLAN	6
Alignment and Governance of Research Data Management.....	6
Research Data Service Management	8

OVERVIEW

Executive Summary

We have completed our research data management (RDM) audit, which covered the period of July through November 2015 to appraise whether adequate controls were in place and in effect for managing research data. To determine the audit scope and objectives, we performed audit planning steps which included obtaining an understanding of strategy and priorities, relevant administrative operations and supporting technology, and conducting a preliminary assessment of risk. We considered systemwide and campus policies, laws and regulations, and criteria established by higher education associations (e.g., DateONE and the Council of Governmental Relations).

At the Berkeley campus, Research Information Technologies (Research IT), in partnership with the Library, is planning and building a program of research data management (RDM) services that will be offered on a consultative basis to the campus research community. Currently, this initiative is centrally funded. The goal of the program is to address current and emerging data management issues, compliance with University and sponsor policy requirements, and reduction of risk associated with the challenges of data stewardship. The services currently being planned and built are a consulting service, an online resource guide, and workshops focused on building a broader base of support for RDM. Implementation of the first phase of these services was just getting started as of the close of our audit, while at the same time Research IT and the Library continues to assess needs, identify opportunities for improvement, prioritize, and plan other further initiatives. The university's California Digital Library (CDL) is also a partner of Research IT and the Library. As their services are offered at the systemwide-level, we did not include them in the scope of our audit procedures.

We observe that the current state of research data management on the campus is decentralized with primary responsibility and accountability assigned to individual researchers. The university does not currently have a complete set of policies pertaining to research data management. Policy and implementation standards or procedures on key topics such as data management planning, active management, preservation, and sharing have not been established. In addition, a system for managing quality assurance in data stewardship, for example prior to completing the pre-award phase and at the close of a project, is not currently part of campus strategy.

We also observed that the partnership between Research IT and the Library to build a community of RDM subject-matter experts and consultants across the campus is in its early stages. We believe this is a positive development. The campus' current investment in bolstering RDM has many potential benefits including promoting compliance with external and sponsor requirements, training on better practices across the entire research data lifecycle, and reduction of risk and increase of efficiency in RDM. However, further efforts are likely needed to address gaps in institutional governance, for managing relationships between the community of service providers and the research community, and for maintaining service catalogs and agreements.

Source and Purpose of the Audit

Each year Audit and Advisory Services prepares an annual service plan of audit and advisory service engagements based on our enterprise risk assessment. Research data management (RDM) was identified as an area of potentially heightened risks associated with the research data stewardship lifecycle. It also aligns with the campus strategic initiative of research innovation. Our audit objective was to appraise whether adequate controls were in place and operating effectively to mitigate risks associated with managing research data. Our assessment was as of December 2015.

Scope of the Audit

To determine the audit scope and objectives, we performed audit planning steps which included obtaining an understanding of management's strategy and priorities, relevant administrative operations and supporting technology, and conducting a preliminary assessment of risk. We considered systemwide and campus policies, laws and regulations, and criteria established by higher education associations (e.g., DateONE and the Council of Governmental Relations) to identify RDM objectives and risks relevant to the campus. We also considered prior audit work but, to our knowledge, this is the only audit on this topic across the system in recent memory.

Systemwide and Campus Policies

Teaching and research records are excluded from the scope of the University Records Management Program, which pertains primarily to administrative records. Section APM-020 (Special Services to Individuals and Organizations) of the Academic Personnel Manual outlines the right of faculty to render service in connection with research undertaken for the benefit of federal, state, industrial or other projects. However, it states "Notebooks and other original records of research are the property of the University." This statement of ownership is also cross-referenced in Chapter 17-200 of the systemwide Contracts and Grants Manual. Neither reference provides further policies or guidance on managing research data. Other examples of relevant policies speak to the systemwide and campus approach to securing information and planning for and responding to disasters or significant disruptions:

- BFB-IS-12 Continuity Planning and Disaster Recovery,
- BFB IS-3 Information Security,
- Data Classification Standard,
- Minimum Security Standards for Networked Devices,
- Minimum Security Standards for Electronic Devices, and
- Committee for Protection of Human Subjects Policies (i.e., Data Security and Data Security Guidelines and Matrix)

Laws and Regulations

Legal and regulatory requirements applicable to research data management include the federal Health Insurance Portability and Accountability Act (HIPAA), the California Committee for the Protection of Human Subject Data Security Requirements, as well as, policies of major federal agencies such as National Institutes of Health and National Science Foundation that sponsor campus research. At the time of our audit, criteria established by higher education associations

and federal requirements were limited, but rapidly evolving. In scoping our audit, we were mindful of this evolving nature of external requirements relative to the current state of systemwide and campus practices.

Background Information

Research IT, in partnership with the Library, is planning and building a program of RDM services that will be offered on a consultative basis to the campus research community. Currently, this initiative is centrally funded. The goals of the program are addressing current and emerging data management issues, compliance with University and sponsor policy requirements, and reduction of risk associated with the challenges of data stewardship. The services currently being planned and built are a consulting service, an online resource guide, and workshops focused on building a broader base of support for RDM. Implementation of the first phase of these services is just starting while, at the same time, Research IT and the Library continues to assess needs, identify opportunities for improvement, prioritize, and plan other further initiatives.

The university's California Digital Library (CDL) is also a partner of Research IT and the Library. The CDL offers the following services:

- DMPTool – a tool for creating, editing, sharing, and saving data management plans,
- Merrit – a repository for managing, archiving, and sharing valuable data,
- EZID – a tool for creating and managing unique long-term identifiers (e.g. metadata and URL locations),
- Dash – a tool for describing, managing, and sharing research data,
- Web Archiving Service – a tool to capture, analyze, archive, and publish websites and documents, and
- eScholarship – an open access scholarly publishing service for papers, journals, books, seminars, and more.

As these services are offered at the systemwide-level, we did not include them in the scope of our audit procedures. However, we did inquire as to their current use on the campus.

Summary Conclusion

We observe that the current state of research data management on the campus is decentralized with primary responsibility and accountability assigned to individual researchers. Policy and implementation standards or procedures on key topics such as data management planning, active management, preservation, and sharing have not been established. In addition, a system for managing quality assurance in data stewardship, for example, prior to completing the pre-award phase and at the close of a project, is not currently part of campus strategy.

We also observed that the partnership between Research IT and the Library to build a community of RDM subject-matter experts and consultants across the campus is in its early stages. We believe this is a positive development. The campus's current investment in bolstering research data management has many potential benefits including promoting compliance with external and sponsor requirements, training on better practices across the entire research data lifecycle, and reduction of risk and increase in efficiency in research data management. However, further efforts are likely needed to address gaps in institutional governance, for managing relationships

between the community of service providers and the research community, and for maintaining service catalogs and agreements. The potential impact can include overinvestment, underinvestment, or selection of the wrong investment alternatives; underutilization of services; harm to future research if the research is not accurately and completely documented, preserved, and discoverable; compliance risks; and risk of data breaches.

SUMMARY OF OBSERVATIONS & MANAGEMENT RESPONSE AND ACTION PLAN

Alignment and Governance of Research Data Management

Observation

We observe that the current state of research data management on the campus is decentralized with primary responsibility and accountability assigned to individual researchers. Policy and implementation standards or procedures on key topics such as data management planning, active management, preservation, and sharing have not been established. In addition, a system for managing quality assurance in data stewardship, for example prior to completing the pre-award phase and at the close of a project, is not currently part of campus strategy.

We also observed that the campus, through a partnership between Research Information Technologies (Research IT) and the Library, is in the process of developing a research data management (RDM) program of services that can be used by the campus research community on a voluntary basis. We believe this is a positive development. However, further additional efforts are likely needed to address gaps in institutional governance.

While institutional governance over this area is just getting started, a system for evaluating, prioritizing, and monitoring campus investments in initiatives, services, assets, and resources in research data management has not been formally defined. In practice, this would start with establishing a common understanding of what value can be realized for purposes of evaluating, selecting, and optimizing investment in the research data management program. In principle, value is both intangible and tangible and examples of ways in which it can be expressed includes financial cost, cost savings, performance (e.g., delivery of fit-for-purpose services, on time delivery of services), alignment with strategy, and preservation of value through appropriate risk taking.

As a campus, the amount of risk we are willing to accept as it relates to research data management is currently only partially defined. It is currently undefined as it relates to

- the quality of data¹,
- the description and documentation of datasets,
- the preservation of data,
- the discoverability of data,
- services that enable and support research data management, and
- compliance with sponsor requirements.

As it relates to protecting confidential information, requirements have been codified in policy through the campus data classification standard as well as the minimum-security standards for network devices and electronic information. However, management has identified potential gaps

¹e.g., Data generated and compiled meets the goals of the project (i.e., relevancy), is consistently represented, accurate, complete, and appropriately documented for ease of interpretation and understandability.

in current security. A task force is being formed to evaluate and assess this situation and make recommendations for changes in existing policy and/or services.

In the absence of a strategic approach to investment in campus support for research data management, the potential impact could include overinvestment, underinvestment, or selection of the wrong investment alternatives; difficulty in determining whether the benefits of investments were realized; and negative impacts to future research if the research is not accurately and completely documented, preserved, and accessible; compliance risks; and risk of data compromises or breaches.

Management Response and Action Plan

Leadership of the campus-wide Research Data Management program is developing a multi-year plan for these areas. Initial steps and their target completion dates are

- The appropriate governance for RDM lies in the campus IT Governance structure. The primary stakeholders for the RDM program (the Chief Information Officer, University Librarian, and Vice Chancellor for Research) are members of the IT Strategy Committee (ITSC), a crucial component of IT Governance. Additionally, all three offices are well represented in the other committees that make up IT Governance, particularly the Research, Teaching and Learning Technologies Committee (RTLTC), which reports to the IT Strategy Committee.
- Further discussions with the Vice Chancellor for Research and his staff will be organized in order to strengthen and formalize the role of the VCR and key members of the Office of the VCR in the RDM program. This will be documented in meeting minutes that will include a draft research data management governance model by August 31, 2016.
- The primary stakeholders will lead the submission of a proposed RDM Governance Model and initial goals to the ITSC and RTLTC by September 30, 2016. This proposal will be documented by the meeting agenda and minutes.
- The proposed governance model will incorporate the formation of the RDM Advisory Group as a subcommittee to the Research, Teaching and Learning Technologies Committee (RTLTC). Comprising high level faculty and RDM program leadership, the RDM Advisory Group will advise RTLTC and the RDM program on priorities for service development, policy work, risk management, and governance. The first meeting of this group will take place before October 31, 2016 and will be documented by the meeting agenda and minutes.
- The RDM program, with guidance from the RDM Advisory Group, will work with policy groups on campus to enhance existing policies related to data stewardship. Specifically, the RDM program will work with Information Security & Policy and the Information Risk Governance Committee to enhance existing policies such as the Data Classification Standard to incorporate more explicit references to research data considerations. A draft revision of this policy will be submitted to the Information Risk Governance Committee by November 30, 2016.

Research Data Service Management

Observation

The RDM core team's current strategy is to establish a campuswide community of consultants that leverages the experience, skills, and capabilities that are already broadly distributed across the campus in both technical and specific academic domains (e.g., humanities, social sciences, and physical sciences).

A challenge with this distributed service model is the complexity of the community of consultants not being under a common reporting structure. The ability to share information between consultants regarding clients' changing expectations, opportunities, issues, risks, and complaints is a key component to the success of this strategy insofar as the client has a seamless service experience across providers. The core team has plans to develop a customer relationship management system. However, it may be more productive and cost-effective to consider a system such as Salesforce that is already available and currently used in parts of the campus.

We identified three campus websites² with service catalogs related to research data management. We observed opportunities to utilize a more standardized approach to selecting and presenting information related to research data management across the providers of RDM services across the campus. The current uncoordinated approach may lead to user confusion, dissatisfaction, and consequently low awareness and utilization of RDM consulting services.

Management Response and Action Plan

Leadership of the campus-wide RDM program is developing a multi-year plan for these areas. Initial steps and their target completion dates are

- The RDM program, with guidance from the RDM Advisory Group, will select the appropriate system for documenting consulting information and sharing information with our stakeholders. This decision will be made by December 15, 2016 and will be documented in meeting minutes of the RDM Advisory Group or RDM program team.
- Regarding the service catalog issue, the RDM program is working with the broader campus service provider community on this campuswide issue. Campus has engaged consultants to develop a service catalog for the campus. By January, 2017, information in the multiple web sites identified in the audit report² will be aligned and coordinated.

² technology.berkeley.edu/it-services, ist.berkeley.edu/services, and researchdata.berkeley.edu