



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

2111 BANCROFT WAY, SUITE 300
MAILCODE #1170
BERKELEY, CALIFORNIA 94720-1170

December 15, 2023

Benjamin Hermalin
Executive Vice Chancellor and Provost

Khira Griscavage
Associate Chancellor

Marc Fisher
Vice Chancellor
Administration

Executive Vice Chancellor and Provost Hermalin, Associate Chancellor Griscavage, and Vice Chancellor Fisher:

We have completed our audit of Data Lifecycle 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 Berkeley IT and the Chancellor's Office 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,

Jaime Jue

Jaime Jue
Director

cc: Interim Chief Information Officer Gabriel Gonzalez
Executive Director Wes Johnson
Vice Provost Lisa García Bedolla
Chief Privacy Officer Scott Seaborn
Senior Vice President and Chief Compliance and Audit Officer Alexander Bustamante
Associate Vice Chancellor and Controller Michael Riley



AUDIT AND ADVISORY SERVICES

Data Lifecycle Management Audit

Project No. 23-773

December 15, 2023

Prepared by:

Robert Asato

Robert Asato
Auditor-in-Charge

Reviewed by:

Jennifer Jones

Jennifer E. Jones
Associate Director

Approved by:

Jaime Jue

Jaime Jue
Director

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

Table of Contents

OVERVIEW	2
Executive Summary	2
Source and Purpose of the Audit	3
Scope of the Audit	3
Background Information.....	3
Summary Conclusion.....	4
SUMMARY OF OBSERVATIONS & MANAGEMENT RESPONSE AND ACTION PLAN	5
Unstructured Data Lifecycle Management Governance.....	5
Campus Administrative Data Storage Technology Roadmap	6
Shared Drive Maintenance.....	7

OVERVIEW

Executive Summary

The purpose of the audit was to evaluate the current maturity of campus business processes, programs, and internal controls related to managing the risks related to the ongoing creation, storage, use, archiving, and destruction of unstructured data.

Our audit scope was primarily focused on data lifecycle management activities associated with unstructured data stored in the campus' file server environment maintained by Berkeley IT. The scope was limited to administrative units and processes and, therefore, research and academic units as well as division or department level platforms were outside the scope of our review.

Unstructured data is created every day by campus users in the form of new or modified files reflecting the creation, extraction, modification, transformation or analysis of institutional information, which can then be further shared through email or other collaboration tools. The importance of effective data management for both structured and unstructured data has been increasing in recent years due to the proliferation of cybersecurity, privacy, and regulatory concerns. On the Berkeley campus, data lifecycle management has been an area of cross-divisional management focus and action through, including but not limited to, information security and privacy-related policies and programs, enhancements to the assessment process for third parties accessing campus data, and the recent establishment of a campus Data Governance Committee. We acknowledge these efforts, as well as the range of competing priorities the campus is facing with respect to the information technology and data landscape, and note certain opportunities to help further evolve and strengthen the campus' longer term position in effectively mitigating data management risk.

While there are various units and positions involved in educating the campus community regarding different aspects of data management, there is currently no overarching formal governance framework or organizational structure to establish and communicate policies and guidance pertaining to data management as a whole. Given the longer term implications, costs, and risks of maintaining large amounts of unstructured institutional data, ensuring an appropriate governance framework and structure is in place warrants additional management attention. In addition, the campus has been proactive in evolving data file sharing and storage platform offerings to meet internal needs and external imperatives, but is currently operating without an overall solutions roadmap to guide current and future decisions around unstructured data storage platforms and programs. We recommend that management begin the process of developing a roadmap for unstructured data storage solutions to ensure ongoing alignment with campus and end-user needs, and to anticipate and be positioned to respond to potential factors outside of campus control.

Lastly, while the responsibility for ongoing access reviews and data management resides with each campus department, we note an opportunity for Berkeley IT management to work with these departments to implement a formal verification process for confirming departmental contacts and user access rights for storage solutions hosted centrally by Berkeley IT.

Source and Purpose of the Audit

Our audit was performed as part of our fiscal year 2023 audit plan. The purpose of the audit was to evaluate the current maturity of campus business processes, programs, and internal controls related to managing the risks related to the ongoing creation, storage, use, archiving, and destruction of unstructured data.

Scope of the Audit

Our audit scope was primarily focused on data lifecycle management activities associated with unstructured data (files not stored in a structured database format) stored in the campus' file server environment, specifically those maintained by Berkeley IT, and was further limited to focus on administrative (versus research or academic) units and processes. Division or department level platforms were outside the scope of our review. Areas assessed include:

- file server provisioning and access management;
- file server/share content management;
- records retention practices; and
- file server decommissioning.

Our internal audit procedures entailed interviews with staff and management from the Berkeley IT Campus Information Technology Experience (CITE) team, as well as a review of information related to the file servers currently under management. We also met with a small number of end-user units to gain an understanding of how and why file servers are used by departments as a preferred storage platform, as well as what department-level processes exist around data creation, access, and disposal. Audit procedures were conducted during the November 2022 to May 2023 timeframe.

Background Information

Unstructured data is created every day by campus users in the form of new or modified files reflecting the creation, extraction, modification, transformation or analysis of institutional information which can then be further shared through email or other collaboration tools. The campus currently offers four primary platforms for sharing and storing unstructured data created by faculty and staff that are hosted either in the cloud or on local servers, including Box, bDrive (Google), CalShare (SharePoint), and file servers (colloquially referred to as shared drives). The campus provides guidance to end-users regarding use cases for certain options, but does not mandate any single solution except in cases where data meeting certain data security classification levels is involved.

The concept of data governance for both structured and unstructured data has been increasingly important in recent years due to the proliferation of cybersecurity, privacy, and regulatory concerns. On the Berkeley campus, there has been a particular emphasis and focus on storage solutions for the research enterprise due to the larger data sets and federal sponsor requirements involved. The campus also stood up a Data Governance Committee in fiscal year 2023, with an initial charge to provide “strategic oversight of the management, access and use of one of the University’s most vital assets—its data.”

Summary Conclusion

Data lifecycle management is an area of critical importance to the campus, given the costs and risks associated with medium- and long-run maintenance of unstructured institutional data. On the Berkeley campus, data lifecycle management has been an area of cross-divisional management focus and action through, including but not limited to, information security and privacy-related policies and programs, enhancements to the assessment process for third parties accessing campus data, and the recent establishment of a Data Governance Committee. We note a further opportunity for management to assess and delineate a more formal and holistic approach to promoting and ensuring appropriate campuswide data management practices for unstructured data across the following areas:

- *Unstructured Data Lifecycle Management Governance*: Ultimately, as creators and custodians of their own unstructured data, each individual employee must act as a steward of their data. While there are various units and positions involved in educating the campus community regarding different aspects of data management, there is currently no overarching formal governance framework or organizational structure to establish and communicate related policies and guidance pertaining to data management as a whole. Given the longer term implications, costs, and risks of maintaining large amounts of unstructured institutional data, ensuring an appropriate governance framework and structure is in place warrants additional management attention. As a collaborative effort, management may wish to determine which unit(s) are best positioned to take the lead in considering and developing programs to strengthen the campus' data lifecycle management position. Additionally, there are various programmatic approaches that management may consider, ranging from general awareness building to tailored training and document management plan development for individual departments or data custodians.
- *Campus Administrative Data Storage Technology Roadmap*: The campus has been proactive in evolving platform offerings to meet internal and external needs, but is currently operating without an overall solutions roadmap to guide current and future decisions around unstructured data storage platforms and programs. We recommend that management begin the process of developing a roadmap for unstructured data storage solutions for the campus to better understand and align with current and future campus data storage behaviors and needs, to manage related costs and risks to the extent possible, and to anticipate and be positioned to respond to potential factors outside of campus control. We also note an opportunity for management to build upon existing guidance to further increase awareness of the available options, including file servers, which are not currently widely publicized, and to lead users to their optimal existing solution in the shorter run.
- *Shared Drive Maintenance*: Lastly, for the one data storage platform in scope of our review (file servers), we note an opportunity for CITE management to work with end-user departments to update contacts and to recommunicate departmental contact responsibilities, and to implement a formal periodic verification process whereby departmental contacts and user access rights to shared drives are confirmed.

SUMMARY OF OBSERVATIONS & MANAGEMENT RESPONSE AND ACTION PLAN

Unstructured Data Lifecycle Management Governance

Observation

Unstructured data is created every day by campus users in the form of new or modified files reflecting the creation, extraction, modification, transformation, or analysis of institutional information which can then be further shared through email or other collaboration tools. Systemwide records retention policies apply only to data that is narrowly defined as an institutional record, but security, privacy, and legal/public records concerns are relevant to all work products containing institutional information. Therefore, ensuring employee awareness of risks, and enabling employee compliance with applicable policies and best practices for data handling is imperative. While some campus departments may already have established protocols and training for staff to ensure appropriate electronic records handling and retention/destruction (especially those accessing and maintaining data covered by HIPAA and FERPA), there is currently limited campuswide guidance, training, and oversight provided.

Given the longer term implications, costs, and risks of maintaining large amounts of unstructured institutional data, this is an area that warrants additional management attention. Ultimately, as creators and custodians of their own unstructured data (even if not deemed institutional records), each individual employee must act as a steward of their data. Accordingly, as a foundational step, management may wish to determine which unit(s) are best positioned to take the lead in establishing and developing programs to strengthen the campus' data lifecycle management position. There are already functions and structures in place that encompass aspects of this work, including the Records Management coordinator role and/or the recently formed campus Data Governance Committee. The Institutional Information Proprietor (IIP) and Unit Information Security Lead (UISL) roles established through the campus IS-3 implementation might also be engaged. Once governance/leadership for these efforts has been defined, there are various programmatic approaches that management may wish to consider, ranging from general awareness building to tailored training and document management plan development for individual departments or data custodians.

Management Response and Action Plan

The campus records management coordinator will take the lead in addressing the audit observation. Planned actions are anticipated to include the following:

- Development of data stewardship guidelines for communication to the campus IIPs. The guidelines are anticipated to encompass aspects of the data stewardship lifecycle, including data identification, definition, retention, and storage. *(Target completion date: July 2024)*
- In addition, the campus Data Governance Committee has already convened a working group to establish guidelines and oversight related to access and use of institutional data. *(Target completion date: December 2024)*
- Finally, as a longer term initiative, the records management coordinator plans to build out campuswide communications regarding records management requirements.

Campus Administrative Data Storage Technology Roadmap

Observation

In recent years, Berkeley IT has expanded its offerings for administrative unstructured data storage solutions. Currently, there are four primary platforms available to campus users that are hosted either in the cloud or on local Berkeley IT-managed servers. We note that each solution entails differing cost/risks and benefits, and addresses different end-user needs. Business needs and priorities surrounding unstructured data storage are continually evolving with respect to vendor pricing structures, the cyber threat landscape, regulatory requirements, and end-user behavior and business needs. Although these factors are known to and considered by management, and changes to storage programs are implemented in response, the campus is currently operating without an overall solutions roadmap to guide current and future decisions around unstructured data storage platforms and programs. We acknowledge the range of competing priorities the campus is facing with respect to its information technology infrastructure and services, and that concerns related to optimizing our storage programs for unstructured data may be of lesser immediate concern. Nonetheless, we recommend that management begin the process of developing a roadmap for unstructured data storage solutions for the campus to better understand and align with current and future campus data storage behaviors and needs, to manage related costs and risks to the extent possible, and to anticipate and be positioned to respond to potential factors outside of campus control.

In addition, Berkeley IT has published high-level guidance regarding appropriate use cases for certain storage options and also offers ad hoc consulting services to units seeking additional information and input regarding which solution might best meet unit needs. We note a further opportunity for management to build upon existing guidance to further increase awareness of the available options, including file servers, which are not currently widely publicized, and to lead users to their optimal existing solution in the shorter run.

Management Response and Action Plan

The development of an administrative data storage technology roadmap is a longer term strategic undertaking that will require the engagement of multiple Berkeley IT units, as well as campus leadership support. Currently, discussions are underway to determine the larger storage strategy that will address the data storage needs of the campus research enterprise.

Storage strategy is shared across several campus departments, including Berkeley IT's Campus IT Experience (CITE) and Campus IT Infrastructure (CITI), and Research Teaching and Learning (RTL). Berkeley IT will look at ways to expand and partner on the guidance provided to the campus community regarding the storage platforms currently available in order to promote the selection/usage of the most appropriate option(s), balancing campus and end-user needs and associated risks. The timing and priority will be determined in FY25 during our annual strategic planning process, contingent on the hire of a new CIO and the potential designation of a new strategic lead.

Shared Drive Maintenance

Observation

One of the four primary data storage platforms offered to the campus are file servers managed by Berkeley IT. These file servers are a legacy solution that only a subset of campus divisions and departments use, and is not a platform that has been broadly promoted to campus as an option in recent years. Currently, there are 19 file servers storing approximately 66.5 terabytes of data across hundreds of file shares for nearly the same number of departments. CITE has established a general governance process for granting file server/share access, and maintains a list of departmental contacts for this purpose. Responsibility for ongoing access reviews and data management resides with each department.

We reviewed records associated with a subset of these file servers/shares and interviewed a small number of designated departmental contacts. Although the file servers/shares we reviewed are still in use, we noted that the vast majority of stored data was created and/or last accessed more than five years ago, and departmental contacts whom we interviewed were not uniformly aware of what data was being maintained. In addition, certain of the named departmental contacts are no longer with the campus and do not appear to have formally transitioned their file server/share responsibilities to a new contact.

We note an opportunity for CITE to work with end-user departments to update contacts and to recommunicate departmental contact responsibilities. We recommend that CITE management implement a formal periodic verification process whereby departmental contacts and user access rights to shared drives are confirmed. This communication to departmental contacts should also include a reminder to review file shares and apply appropriate record retention policies for data no longer in use and to delete data, as appropriate.¹

Management Response and Action Plan

CITE agrees with the findings of this audit, and will be reviewing our full suite of storage offerings, including Google, Box, and file shares, to determine a new strategy to address the gaps across the board. Due to limited resources, lack of available storage options for some campus needs, and change management requirements, we are unable to commit to a specific date for this work at this time. We are in the early stage of implementing access control via CalGroups. This will enable a more intuitive method for units to review and manage access for their file shares.

In the interim, IT Client Services (ITCS) is developing a plan to update our contact lists for file shares in the next year, including confirming the current ownership of the individual file shares. *(Target completion date: July 2024)*

¹ Although outside the scope of our audit, we note that this condition likely is relevant to other campus storage and file sharing sites and management may wish to consider developing a common verification process across platforms.