UNIVERSITY
OF
CALIFORNIA

Recommendations to Protect UC's Digital Research Data

Cyber-Risk Governance Committee

November 23, 2020

Contents

1. Executive Summary

At the recommendation of the Cyber-Risk Governance Committee (CRGC), in July 2020 President Napolitano established a Cyber-Risk Working Group (CRWG) in response to the June 2020 ransomware incident at UCSF. The CRWG was chartered in August and began work in September to gather information and make recommendations for better managing the security of digital research data and for protecting UC's research enterprise.

Working group members examined the structural, technical, financial, and cultural issues across UC that impact UC's ability to effectively protect digital research data. The working group focused on pragmatic and balanced actions that will elevate UC's security posture for digital research data and lower the risk of data being destroyed or made unavailable.

The recommendations reflect the people, process, and technology efforts required to holistically address digital research data security and may be executed in parallel. It is expected that the CRGC will provide coordination oversight.

The CRGC endorses the working group's recommendations and recommends systemwide adoption. The recommendations, more fully described in the Recommendations section below, are for UC to:

1. Establish location-based research data protection workgroups,
2. Develop awareness initiatives to enable workplace environment change, and
3. Provide a scalable data backup service for all UC researchers.

The CRGC believes these recommendations will make a positive difference in securing digital research data and minimizing threats. However, despite its current fiscal challenges, the University will have to develop an approach to fund these recommendations, so that lack of resources is never the reason the University's digital research data are not protected. The reduction of risk and protection of digital research data must be ongoing and must evolve as circumstances change. In other words, the University must remain vigilant and committed to protecting its valuable assets.

The committee formally requests that the President endorse these recommendations and charge the CRGC to take the necessary actions to ensure their implementation. UC will need to balance the urgency of protecting digital research data with the cost to UC and the impact of these efforts upon researchers. The CRGC looks forward to the opportunity to meet with the President to discuss this report, address any questions and concerns, including potential success metrics, and begin planning the implementation of these recommendations and addressing resource considerations.

2. Introduction

On July 28, 2020, Interim Vice President and Chief Information Officer Mark Cianca and Vice President for Research & Innovation Theresa Maldonado wrote to President Napolitano to provide an analysis by the Cyber-Risk Governance Committee (CRGC) of a ransomware incident at UCSF involving digital research data. The circumstances surrounding this incident raised significant concerns within the research enterprise and across UC, particularly regarding the security of digital research data and the availability of technology for the research community to protect data.

President Napolitano endorsed the actions recommended by the CRGC, including establishment of the Cyber-Risk Working Group. To form the group, VP Cianca and VP Maldonado solicited volunteers from the CRGC, the UC Vice Chancellors for Research, the Academic Senate, and other subject matter expert groups, such as the California Digital Library (CDL) and the Lawrence Livermore National Laboratory (LLNL). VP Cianca and VP Maldonado issued a charge letter to the working group members to examine and address the issues germane to digital research data protection and make recommendations to the President to strengthen cybersecurity in the research enterprise. They asked that the working group consider the following:

1. Identifying the available services, resources, and best practices for securing research data; identifying gaps and proposing solutions to address high-priority items,
2. Identifying structural, technical, financial, and cultural challenges that impact faculty, staff, postdoctoral, graduate, and undergraduate researchers, and proposing plans to address them,
3. Working with the Academic Senate on developing and delivering guidance to researchers on how to apply appropriate security controls for their environments,
4. Proposing a system level research data lifecycle management program to be adapted to the campus level, and
5. Proposing a framework for Vice Chancellors for Research (VCRs) to establish an ongoing workgroup of researchers at each location to facilitate the sharing of ideas, challenges and successes in managing cyber risk, as well as regular reporting to the CRGC on the status of these efforts.

This report is the result of the working group's efforts and is intended as a plan to address the structural, technical, financial, and cultural challenges facing UC in its efforts to better manage cyber risk in the research enterprise.

3. Working Group Members

The Cyber-Risk Working Group was comprised of the following members:

Name	Title
Pamela Miller	**Interim Assistant Vice Chancellor, Research Administration and Compliance, UCB**
Kenneth Lutz	Interim Director of Research IT, UCB
Chris Hoffman	Associate Director, IT Research, UCB
Prasant Mohapatra	Vice Chancellor of Research, UCD
Phil Papadopoulos	Director of the UCI Research Cyberinfrastructure Center, UCI
Susan Cochran	Academic Computer and Communications Vice Chair, UCLA
David Shaw	Chief Information Security Officer, UCLA
Marjorie Zatz	Interim Vice Chancellor for Research and Economic Development, UCM
Charlie Eaton	UCM, Professor of Sociology, UCM
Lisa Yeo	Assistant Professor of Management of Complex Systems, UCM
Mary Gauvain	Academic Senate Chair, UCR
Rodolfo H. Torres	Vice Chancellor for Research and Economic Development, UCR
Joe Incandela	Vice Chancellor for Research, UCSB
Jean Jones	Assistant Vice Chancellor for Research, UCSB
Scott Brandt	Vice Chancellor for Research, UCSC
Robert Horwitz	Academic Senate Vice-Chair, UCSD
David Robinowitz	Academic Computing and Communications Chair, UCSF
Joe Bengfort	Chief Information Officer, UCSF
John Chodacki	California Digital Library, Director of the UC Curation Center

Name	Title
Gregory Herweg	Chief Technology Officer, LLNL
Matthew Myrick	Deputy Chief Information Security Officer, LLNL
Mark Cianca	Interim Vice President and Chief Information Officer, UCOP
Theresa Maldonado	Vice President for Research & Innovation, UCOP

4. Methodology

The working group assembled five sub-groups, each with a lead, to examine the five considerations outlined in the charge letter (see the introduction to this report): digital research data, challenges, faculty awareness and engagement, emerging practices, and the research data lifecycle.

Each sub-group researched the issues, which included reviewing written materials and interviewing colleagues, and developed a set of findings and recommendations for that area. The findings were brought back to and discussed among the larger working group, and ultimately contributed to the working group's overall recommendations for the protection of digital research data, as presented in this report.

5. Key Findings

Digital Research Data

Digital research data at UC represent key intellectual property for the investigator, the research group, the department, the campus, and the University system as a whole. Traditionally, security is defined as ensuring the confidentiality, integrity, and availability of the data. A great deal of attention is focused on *securing the access* to data that must be closely held and protected. This type of protection seems to be well implemented for data with statutory protection requirements, such as health (HIPAA) and student (FERPA) data. In these areas, there is understanding of the risks to this type of data; and risk-based measures, such as inventories, vulnerability testing, and cybersecurity training, exist as standard practice.

However, digital research data comprise a vastly more expansive domain than data that must be protected by statute (which is covered by protection level 4, as defined by UC's information security policy IS-3) or data considered sensitive (covered by protection level 3). There are significant amounts of “public” or “internal” digital research data within UC. It is voluminous by comparison with the data protected by statute; and managing the scale of these vast amounts of data is a significant challenge, as individual datasets can range over twelve orders of magnitude (bytes vs. petabytes). While many non IT professionals believe data security involves only encryption and rigorous access control (*confidentiality*), in fact, data *integrity* and *availability* are just as critical to the protection of digital research data.

While digital research data have become pervasive throughout every branch of research at UC, a large fraction of investigators, their students, and colleagues often remain unaware of where and how their data are stored or the risks to which their data are susceptible. UC is somewhat blind to the risks to data that do not require rigorous access and encryption protections. Cost and scale contribute to this situation but are not the only obstacles. Of particular concern is the issue that data may be unknowingly and incidentally modified. Such modification (through storage system failure, errant day processing, human error, and more), if undetected, can have potentially catastrophic effects on scientific analysis and conclusions.

Although many researchers intellectually understand the possibility of large-scale, complete data loss, given the current widespread practice of inadequate data replication (or backup), it is apparent that many do not act themselves or direct their research programs to perform regular, offsite, copy of data. Too many, it seems, fall into the trap of thinking, “If nothing has happened in the past, it will not happen in the future.” Further, data backup and validating the integrity of the backups can be time-consuming and potentially unremunerated by sponsored research projects, especially once those projects have completed their funding cycle. But data do not have to be under attack by hackers to be at risk. And at UC, far too much research data are, indeed, at risk.

Examples of this risk can be seen in recent incidents, such as:

- **UCSF**: The June 2020 ransomware attacks severely disabled a department and led to the payment of $1.14M in ransom.
- **UCSC**: The August 2020 wildfires threatened the Santa Cruz campus. While 500TB of data were copied to remote sites, other data that could not be moved were covered with tarps to protect against water damage. Fortunately, no data were lost.
- **UCI**: In September 2019, a burst water pipe flooded a portion of the School of Medicine. Several computers were destroyed and a large trove of digital research data was made unavailable. These data were needed for a multimillion dollar NIH grant submission but the NIH refused to allow a late submission, stating the data were unavailable due to negligence (because there was no second copy) rather than a natural disaster.

- **UCSD**: June - September 2020, a re-forecast of an atmospheric model was required. The data were stored on magnetic tape with no redundancy. The previous forecast model output had been deleted due to space concerns, and when recovery of the backup was attempted, there was a physical failure. The input data set was lost, making repeated simulations impossible, and therefore preventing key scientific insights that would have improved models and forecasts.

Challenges

In examining the issue of digital research data protection, several issues come to light. First, the decentralized and highly independent nature of the research enterprise, while enabling adept and innovative research, has several downstream consequences, including:

- Uneven understanding by practitioners of the risk to digital research data
- Varied data management practices for the protection of research data
- Gaps in coverage of backup services for both small- and large-scale datasets
- Gaps in communication and collaboration among faculty, researchers, and the administration

Second, these consequences are often lost in the operational shuffle as locations and researchers do their work. While some centralized backup solutions are offered, many labs and individual researchers must fend for themselves, and often turn to vulnerable, external hard drive storage. Given there is no one-size-fits-all approach, there is great need for consultation between researchers and IT organizations. However, other operational security activities, such as encryption, threat detection, vulnerability scanning, and incident response overwhelm many IT organizations, leaving less time for higher-order, consultative interaction. In some cases, there are readily available solutions, but these require that the researchers not only be made aware of them, but also be provided the support to participate effectively in protecting their data.

Third, discussions about researcher needs, goals, and priorities are often overshadowed by the ever-growing external regulations and internal policies impacting research, combined with the ever-present struggle for resources and funding. Yet UC needs to foster a workplace environment that encourages collaborative dialogue about data protection among faculty, researchers, and the administration. While it is unreasonable to think UC can eliminate all the risks to research data, or that there will not always be many competing priorities for funding, a collegial work environment among faculty, researchers, and IT staff is critical to enable the UC community to work together to protect digital research data.

Faculty Awareness and Engagement

UC faculty and researchers are keenly aware of the role of technology in their work. Technology is an essential tool for them to accomplish their research, which brings public recognition, financial support, and prestige to the University. They are also aware of the University's increasing dependence on technology, as evidenced by the pandemic-driven move to remote instruction and collaboration.

Faculty and researchers are affected not only by evolving and dynamic cybersecurity threats, but also by the growing regulatory requirements related to cybersecurity. They need support to navigate these challenges so they may successfully continue their work. The exponential growth of digital research data across all disciplines requires that UC re-direct resources (time, funding, work prioritization, etc.) toward data protection and, where expertise already exists, undertake the necessary organizational realignment and process redefinition to support researchers in protecting data.

The faculty would benefit from bi-directional education and outreach activities, conducted in conjunction with the Academic Senate, such as:

- Frequent and clear communication about the risks to research data, as well as solutions to address those risks
- Consultation and guidance on the UC Minimum Security Standard
- Backup and recovery services for their digital research data

Emerging Practices

As incidents of ransomware and intellectual property theft rise, some UC locations are exploring and implementing new approaches to protecting their cyber-infrastructure. For example, UCSD recently implemented a model to help reduce risk for new sponsored research. Given the rise in ransomware, increasing regulations, and the shift by the Department of Defense to the Cybersecurity Maturity Model Certification (CMMC), UCSD defined and internally published a baseline standard based on CMMC Level 1. This standard ensures common security controls (e.g., anti-malware software, regular patching) and record-keeping are deployed, and provides visibility for early detection of attacks and compromises, which is critical for response and remediation.

Labs are required to self-certify compliance with the UC San Diego Baseline Assessment, and this ensures local accountability is identified for labs and distributed IT assets. Local IT teams and labs, working with central IT, can improve their cybersecurity position, which UCSD believes could contribute to a competitive advantage for researchers on some grant applications. By shifting the culture toward proactive cybersecurity management, periodic assessment, and continuous improvement, UCSD is taking significant steps to prepare for federal imposition of CMMC or similar standards.

Research Data Lifecycle

While examining the topic of the research data lifecycle, and its relevance to the protection of digital research data, the working group found that effective stewardship of research data requires a complex network of coordinated infrastructure, services, policies, and expertise. Whether done systemwide or at the campus level, or some combination of the two, this coordinated network must integrate research and academic administration, information technology, the library, and security to provide seamless, just-in-time support throughout the research lifecycle, from inception to completion. The successful implementation of this coordinated network benefits the faculty, individual campuses, and the entire University. It

- Enables security vulnerability and risk assessment and mitigation,
- Protects intellectual property,
- Facilitates data management, use, and reuse,
- Improves the success rate for publication and extra-mural funding opportunities, and
- Fosters and streamlines collaboration.

The lifecycle of digital research data is only one part of the *larger research process*, which includes:

- Planning and research
- Development and execution
- Synthesis and publication
- Conclusion of the project

Each phase of the larger research process has research support needs – infrastructure, services, policies, and expertise. It is essential that UC provide for these support needs in each phase of a research project. When considered in the larger context of a research project, the digital research data lifecycle model should focus on the following:

- **Planning for data creation and/or re-use**: In this phase, researchers should design plans for generating and/or managing data to streamline future uses by themselves and/or others.
- **Organizing research data**: Researchers should organize data clearly so that others can navigate, understand, and use it without them being present.
- **Saving and backing up**: Researchers should save their data in a manner and location designed to clearly articulate the possible opportunities for re-use by themselves and others.
- **Preparing for analysis**: Researchers should prepare data in such a way as to facilitate use by both themselves and others in the future.

- **Analyzing data**: Researchers should ensure that the specifics of their analysis workflow and decision-making process is documented and executed.
- **Sharing and publishing**: Researchers should efficiently share their data to support research findings in accordance with applicable policies and guidelines.
- **Long-term preservation**: Researchers should follow pre-established policies and guidelines for long-term preservation and/or re-use.

6. Guiding Principles

The Cyber-Risk Working Group (CRWG) developed the following five principles to guide UC in planning how best to protect digital research data:

1. **Digital research data is essential to UC**. Digital research data is an essential component of the UC research enterprise and results in discoveries, innovations, and creation of intellectual property. UC researchers, faculty, students, and staff require digital research data to conduct their work and achieve the many successes for which UC is renowned. For this simple reason, digital research data requires protection.
2. **Data protection is a shared responsibility**. Everyone at UC, at all campuses and health systems, has a responsibility to protect UC's digital research data. However, much of the data is at risk of loss or corruption due to:
    - Risks that are not fully appreciated,
    - Insufficient University-supported resources to enable everyone to reduce risk in a practical way, and
    - An increase in digital data that has outpaced the entire community's ability to react adequately and protect it.
3. **Losses can be significant**. Failure to take appropriate steps that dramatically reduce risk can lead to the loss of reputation/credibility, the missed opportunity for new avenues of research, and financial losses in the event of a breach or ransomware.
4. **UC should set data protection goals**. UC should take a goal-oriented approach to the threat, starting with ensuring that, in three years, the preponderance of digital research data held across all UC research labs and facilities will be at low risk for loss or corruption and its protection will be considered an integral part of long-term data lifecycle management.
5. **UC will be a leader**. UC should lead the higher-education research community by establishing the protection of digital research data as a paramount concern and making it foundational to the research enterprise.

7. Recommendations

The CRWG recommends the President take the following actions to protect digital research data. These recommendations support the guiding principles outlined above and are an initial and important step toward improving digital research data protection at UC. Resources will be required to fully support these recommendations, and resource considerations are noted with each recommendation.

Recommendation 1 – Establish Location-Based Research Data Protection Workgroups

The Vice Chancellor for Research (VCR) and Chief Information Officer (CIO) at each location should establish a standing Research Data Protection Workgroup (RDPW) to:

- Facilitate the communication of best practices for protecting digital research data throughout the research data lifecycle,
- Develop and continually enhance awareness materials, templates for researchers, and associated processes and procedures to protect digital research data,
- Communicate to researchers the available resources and services for protecting their digital research data, and
- Enhance the understanding of researcher needs; and share ideas, challenges, and successes about cyber-risk management.

The location workgroups should include the following roles:

- Vice Chancellor for Research
- Cyber-Risk Responsible Executive
- Academic Senate / Location Academic Senate IT representative(s)
- Research IT / Cyberinfrastructure representative(s)
- Library – Research Data and Digital Preservation representative(s)
- Sponsored Programs representative(s)
- Institutional Review Board representative(s)
- Location Chief Information Officer
- Location Chief Information Security Officer
- Export Control Officer

The Vice President for Research should propose a method to enable the sharing of information (including emerging practices, such as the UCSD example noted above) and materials across locations and workgroups.

The workgroups should meet locally on a quarterly basis, at a minimum, and consider meeting monthly for the first six months to jumpstart their work. The workgroups should collectively report twice annually to the CRGC on accomplishments and planned activities.

Resource considerations. There will be a cost to administer these workgroups. The CRGC recommends that the President establish an expectation with chancellors that they incorporate this effort into their planning activities. Given the criticality of the problem, seed funding is required to incentivize work on this recommendation and new resources will be necessary to support a long-term solution.

Recommendation 2 – Develop Awareness Initiatives to Enable Workplace Environment Change

At each location, the VCR, in conjunction with the Provost's Office and Academic Senate, should establish a means to raise awareness about and deliver information from the local Research Data Protection Workgroup to the research community, including faculty, visiting scientists, post docs, and graduate students. The awareness initiative should include:

- Communication about the availability of resources for consultation, solutions for research data protection, and good data protection practices, and
- Examples of and/or templates for data management plans that meet common funding agency requirements, as well unfunded research best practices, to ensure a reasonable data protection approach is documented and completed to support campus compliance efforts.

Because the specialized function to provide educational materials and consultation for protecting research data does not currently exist within the Provost's Office, the CRGC recommends that the Research Data Protection Workgroup develop the required awareness materials, templates for researchers, and associated processes and procedures, and continually enhance them based on researcher needs. The materials should include the established

- UC Minimum Security Standard (https://security.ucop.edu/policies/security-controls-everyone-all-devices.html), and the
- Classification of Information and IT Resources in the UC Information Security Policy, IS-3 (https://security.ucop.edu/policies/institutional-information-and-it-resource-classification.html).

These standards and classifications levels outline required protection practices beyond just backups, such as anti-malware, patching, passwords, and encryption.

Resource considerations. There will be a cost to create and distribute these materials and to establish and implement these processes. The CRGC recommends that the President establish an expectation with Chancellors that they incorporate this effort into their planning activities. Given the criticality of the problem, seed funding is required to incentivize work on this recommendation and new resources will be necessary to support a long-term solution.

Recommendation 3 – Provide a Scalable Data Backup Service for All UC Researchers

UC should centrally fund and provide a uniform backup and recovery solution for each location to administer for its researchers. The solution should be available to all faculty and researchers at no direct cost for all their research data, regardless of IS-3 protection level, including free-standing documents, data files, or application code. It should be implemented in a tiered manner to enable flexibility based on specific need and risk. The solution must:

- Be simple to set up and use ("set and forget") and work on multiple platforms,
- Be integrated into UC's Single Sign-on (Shibboleth), and support multi-factor authentication (or the equivalent),
- Keep the backup research data logically and physically separate from the operational research data, and
- Offer concierge-style support to assist researchers in the backup and recovery of their data.

Resource considerations. Given the cost of providing this solution at a scale to support all UC researchers, and given the benefit of leveraging UC's size for favorable pricing, the CRGC recommends that UC Procurement conduct a systemwide RFP, beginning in early 2021 and to be completed within 12 months, to select a solution that all locations would utilize. The locations' research IT/cyberinfrastructure representatives, CIO, VCR, and Academic Senate representative should direct and support development of the RFP. This recommendation will have resource requirements – both the solution itself and the resources to implement it. Costs will need to be developed as part of the RFP process. A funding model will need to be established, with the traditional campus assessment model being a suggested approach.

8. Conclusion

The protection of digital research data is critical to the continued success of the UC mission, and the increasing threat of ransomware and other disruptive events requires that all members of the University community take action to safeguard the data. The CRGC believes that the recommendations in this report will make a positive difference in securing digital research data and minimizing threats. However, these are only initial steps. The reduction of risk and protection of digital research data must be ongoing and must evolve as circumstances change. In other words, the university must remain vigilant and committed to protecting its valuable assets.

The committee formally requests that the President

- Endorse these recommendations,
- Charge the CRGC to take the necessary actions to ensure their implementation, and
- Consider seed funding to incentivize implementation of the recommendations.

The CRGC recognizes that resources will be required to enact these recommendations. However, the risk is high, the need is significant, and the costs will be new. The University will have to develop an approach to fund these recommendations, despite the current fiscal challenges, so that lack of resources is never the reason the University's digital research data are not protected.

The CRGC looks forward to the opportunity to meet with the President to discuss this report, and address any questions and concerns, including high-level success metrics, and to begin planning implementation of these recommendations and addressing resource considerations.

In closing, the committee would like to thank the members of the working group for their significant and thoughtful contributions to this report, especially in light of the tight timeline, driven by the critical event that brought this topic to the attention of the Regents and the President. The collaboration and focus demonstrated by the working group throughout this process illustrates how the University community unites to protect the shared mission.