

Larry D. Conrad, University of North Carolina at Chapel Hill  
Ruth Marinshaw, University of North Carolina at Chapel Hill  
Stan Waddell, University of North Carolina at Chapel Hill

## Overview

The effective protection and management of research data has become a hot topic in U.S. higher education. For example, the NSF Office of Cyber Infrastructure Campus Bridging Task Force sponsored a workshop in the fall of 2010 where this was a focus of the discussion.<sup>1</sup> Princeton University also hosted an NSF-funded workshop on Research Data Lifecycle Management in July 2011.<sup>2</sup> In particular, funding agencies increasingly require data management plans as part of grant submittals, and research offices are being asked to certify the security of research data generated by grant activity. Heretofore, the context for data management and information security activities and initiatives in higher education largely focused on the “enterprise” (administrative) data of the institution, not those data generated by research activities. Consequently, researchers and information technologists alike are in general unaccustomed to this attention to and focus on the protection of research data.

IT professionals need to be aware that many academic research endeavors include the collection, analysis, and/or storage of sensitive data, the integrity, confidentiality, and availability of which must be asserted and demonstrated. In many cases, the security of sensitive information gathered in the conduct of research is required by law (see the “Where to Learn More” section at the end of this bulletin). Federal regulations such as the Health Insurance Portability and Accountability Act (HIPAA), the Federal Information Security Management Act (FISMA), and proposed changes to the Common Rule highlight the importance of protecting research participants’ identities, research responses, and associated information.

Protecting information in higher education computing environments is challenging in many regards. Campus networks are typically home to thousands of networked devices, control over which is dispersed among central IT units, distributed IT staff, and individuals without a specific IT “home” (for example, students). Openness is often considered a requirement for resources in the academic setting, leading to university networks and systems designed to deliver fast, efficient, and user-friendly services with minimum administrative burden for end users. As such, college and university networks and systems are primary targets for cyber attacks. You might say we are a “destination resort” for the worldwide hacker community. Successful attacks can result in large and costly remediation, and investigation workloads have significant impacts on productivity and can endanger a researcher’s ability to continue their research. Indeed, the Ponemon Institute’s 2009 Fifth Annual US Cost of Data Breach survey found the average cost per compromised record per breach to be \$204.<sup>3</sup> This figure does not include the reputational costs associated with breaches. One study reported that education-related organizations accounted for 31% of all breaches logged by privacyrights.org from 2005 through 2008. More than 12.4 million students and their records were

potentially compromised in 324 security events. Higher education was involved in 79% of these education-related incidents.<sup>4</sup>

Comparing notes with other research-intensive institutions, we find that research information security issues are being raised quite broadly and are not unique to any particular research institution. This research bulletin discusses an overarching approach by which campus IT solutions can be architected and deployed in such a way as to provide adequate management of research data assets without hindering the research process.

## Highlights

The National Academies' report "Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age" begins with the observation that the rapidity and scope of changes in computing, storage, instrument, and networking technologies present both opportunities for research to advance and challenges for ensuring data confidentiality, integrity, and availability.<sup>5</sup> This intersection of growing amounts of research data, open campus networks, distributed systems, distributed and sometimes unarticulated responsibilities for data protection and management, and increased regulatory and legal requirements around data security has brought CIOs, information security officers, research officers, research computing leaders, librarians, and researchers to the same table. At the University of North Carolina at Chapel Hill, we are addressing overall information security through both policy and technical controls. For research data specifically, we have augmented these efforts with additional services and initiatives designed to ensure that researchers have information about, options for, and assistance in data protection and data management.

## Establishing a Policy Base

Over the course of the 2010–2011 fiscal year, the University of North Carolina at Chapel Hill implemented 10 new information security policies with implications for both research and non-research data. With nearly 1,000 staff across the university involved in IT support, and with faculty and students engaged in diverse and distributed activities, using a broad range of information technology systems and services in different ways, it was essential to establish a policy base from which to operate in order to protect campus assets.

The regulations put into place comprised policies related to:

- Overall Information Security
- Information Security Standards
- General User Passwords
- Systems and Applications Administrator Passwords
- Transmission of Sensitive Information
- Security Liaisons
- Vulnerability Management
- Incident Management
- Data Governance
- E-mail

These policies, based on information security best practices, establish a multifaceted approach to protecting the university's IT assets. They were vetted with campus administrative and IT leadership, as well as with faculty and the university's legal team, prior to adoption. Because of the far-reaching scope of the policies, it was essential to collect feedback on them from a broad set of constituencies, while also using the policy discussions as opportunities to develop a shared understanding of the importance of information security. From this policy base, a variety of operational security programs and procedures have been implemented. Each implementation was vetted and commented on by key stakeholders from a wide cross-section of the campus community, including researchers. (We discuss efforts to manage the security of research data in the Research Data Services section below.) One of the first initiatives was the information security liaison group, which was used as—and continues to be—a sounding board for initiatives. The information security liaisons and other groups also help develop the implementation strategies for processes like the system administration initiative described below. Various awareness and outreach methods, such as e-mail, web content, and departmental meetings, were used to communicate initiatives to the campus community.

## **It Takes a Village: Distributed Security Responsibilities**

In a university setting, it is not uncommon for the CIO to be held responsible for overall campus IT security. While UNC's leadership has conferred broad authority and responsibility for information security upon the vice chancellor for Information Technology and CIO, the strategy adopted to implement campus information security policies is one of collaboration. The central information security officer (ISO) and his team organize and lead efforts, but they do so in partnership with research teams, academic departments, and administrative units. The CIO has communicated broadly to the campus community that information security is a shared responsibility, requiring the participation, cooperation, and involvement of campus central IT, distributed IT, faculty, students, and staff. This communication was accomplished through memos, e-mails, presentations for selected groups, and articles in the student newspaper.

Key operational components of the UNC campus information security program include a Security Liaison program and a System Administration Initiative. The Security Liaison program requires each university department (including research units, such as centers and institutes) to identify one person to serve as a liaison between the campus ISO and the department. Meeting monthly, the liaisons help with reporting security incidents, clarifying policy, communicating information from the ISO to individuals in the department, implementing policy, and providing input on information security issues and policies. The Security Liaison program has fostered a sense of community among the individuals serving in these roles, and it has reinforced the extent to which security programs benefit from the active participation of those affected by policy, rather than a top-down, "do this or else ..." approach to information security. The liaisons are not necessarily computer support personnel; rather, the policy targets people with sufficient authority in campus units to effect change.

A new program developed to help ensure the security of campus systems, the System Administration Initiative (SAI), was designed to monitor and evaluate the effectiveness of systems administration across the university for researchers and non-researchers alike. As noted earlier, while UNC has a substantial number of centrally managed technology resources, there is an even greater number of distributed systems—especially from a researcher's perspective. It is not uncommon for individual lab groups or research teams to have an assortment of computing systems and storage devices they manage themselves, with administrative responsibility transferring from one graduate student cohort to the next. This is not necessarily a problem, but it can be, especially if sensitive data are involved in

the research efforts and if the administrators do not have the appropriate knowledge, training, or skills to verify the data's security.

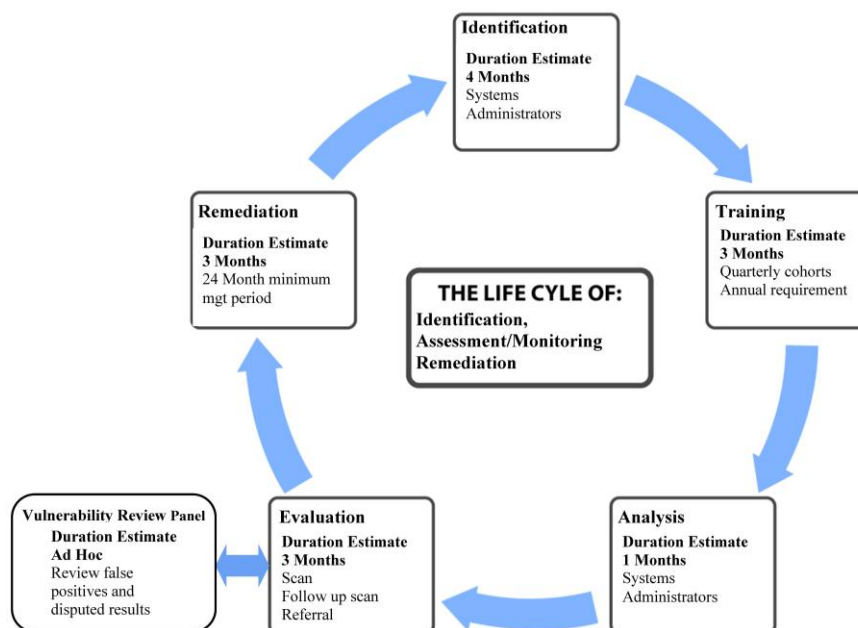
At a high level, the SAI comprises the following elements:

- System administrator identification
- Server identification for systems containing sensitive data
- System administrator training, focusing on expectations about information security and responsibilities in light of those expectations
- Ongoing scanning, monitoring, and assessment of servers to validate proper maintenance and absence of persistent security weaknesses
- Fee-based system administration services for servers identified as having persistent security weaknesses, with costs charged against the campus organizational unit involved

UNC is in the process of creating an assessment framework (see Figure 1) for the continuous monitoring and assessment of servers with sensitive data:

- System administrator identified
- Timely installation of system and application patches according to campus policy
- Access control methods applied correctly
- Adheres to principles of least function (i.e., has running only the services that support the primary function of the device)
- System administrator completed required campus system administration training
- Runs a current version of an antivirus tool
- Encrypted if mobile and contains sensitive information

**Figure 1. System Administration Initiative Process Life Cycle Chart**



Where resources are identified as a primary cause for inadequate system administration, department heads are advised to work with their departmental support personnel to devise a plan for remediating deficiencies and/or providing additional resources to correct the security issues.

The policies, SAI, and the appointment of security liaisons are part of an overarching security plan for the University of North Carolina at Chapel Hill. The overall plan also includes initiatives such as whole disk encryption of laptop computers that store sensitive information, implementation of additional firewall protection, increased focus on the security of research data and servers, and enhanced user outreach, among others. The CIO's office and the Office of Information Security worked directly with IT leaders across the campus to architect the approach. All involved realized that there are no ways to guarantee that the university will be safe from attacks, but there was consensus that the outlined measures provide an effective way of increasing security at UNC Chapel Hill.

### **Research Data Services**

The information security policies and programs detailed above are applicable to all UNC campus IT systems and users, not just those involving research data. For example, research teams are represented in the Security Liaison program whether they are in a separate center/institute or covered by their respective school/department. Also, we think the greatest benefit from the SAI program will be improving security for servers supporting research. For research data in particular, central IT and other campus units have developed and offer a spectrum of services to assist researchers in data management, data protection, and working with sensitive data. These services are intended to help faculty from the grant preparation stage through data archiving at the end of a project.

The UNC University Library and the Odum Institute (a multidisciplinary social science university institute), in collaboration with other campus units such as Research Computing and the Renaissance Computing Institute (also a multidisciplinary, multi-institutional research institute), have developed a Research Data Toolkit, consisting of a set of resources to help researchers write a data management plan. Increasingly, funding agencies require data management plans as part of the proposal project, and some Institutional Review Boards also require a demonstration of researchers' abilities to protect research data throughout its life cycle. The toolkit includes sample data management plans, references to campus resources available for data management assistance, and special requirements related to protecting confidential data.

Research Computing, a division of the central IT organization (Information Technology Services, or ITS), has also developed a set of computational and storage services for use when processing and working with sensitive research data. Depending on the sensitivity of the research data, Research Computing can offer virtualized or stand-alone servers, either Windows- or Linux-based, with accompanying storage and system administration. Some projects might require a firewalled stand-alone environment that meets FISMA standards; others will need a shared virtualized environment with firewalling, restricted system services, and data loss prevention software. The latter services are being deployed for clinical research, having been prototyped by the Renaissance Computing Institute and the TraCS Institute at UNC. (The TraCS Institute leads North Carolina's activities associated with the NIH Clinical and Translational Science Award.) Secured data transfer services are also available, as are customized secured Linux servers with multiple levels of data protection to accommodate research teams in which some members require full file access and some are allowed limited file access. Research Computing works closely with the University's HIPAA compliance officer, the ISO, campus networking, distributed IT staff, and individual researchers to ensure that service offerings meet the legal and regulatory requirements of particular research projects.

While “official” responsibility for information security was delegated to the CIO, it is clear that many other stakeholders share responsibility for protecting research assets. At UNC Chapel Hill, the Chief Research Officer and the CIO have collaborated with the provost to form a task force to look at issues of data responsibility for our institution (see the section below, What It Means to Higher Education); that task force is scheduled to report out in early 2012. The recommendations of that task force, as well as those of other governance groups, will continue to inform and shape the information security ecosystem at UNC. Successful protection and management of research data can only be accomplished through collaborative engagements, broad discussions, and widespread consensus around the importance of the value of data security to fulfillment of the university’s research mission.

## Security Is Not a Solved Problem

Information security issues related to research data are not unique to UNC. Indeed, as noted earlier, we find that many of our peer institutions are struggling with these issues.

Larry Conrad presented on this at the EDUCAUSE Enterprise 2011<sup>6</sup> and Dartmouth Securing the e-Campus 2011<sup>7</sup> conferences in recent months. The former tends to be attended by CIOs and enterprise IT managers, while the latter is attended by information security officers. Based on the strategies we had evaluated at UNC, Conrad conducted an informal poll of the attendees in each session about which strategies they thought might be best for improving research data security. He hoped to get feedback on whether we were overlooking or misprioritizing something important. Attendees were asked to vote for what they considered the single most important strategy. The poll results appear in Table 1. (The last strategy on, “Increased training for researchers,” was suggested at the Enterprise conference and subsequently added to the poll for the Dartmouth conference, which is why that option shows “n/a” for the Enterprise conference.) While anecdotal and not at all definitive, these two polls yielded surprisingly similar results, given the different communities involved. The one obvious exception is the perceived value of including the research/compliance offices—this perhaps suggests some skepticism in this regard among information security professionals. At any rate, we believe these limited data illustrate an understanding among the attendees that a multifaceted approach like the one we are using at UNC is needed to improve security for research data. There is no one-size-fits-all solution.

**Table 1. Which Strategy Holds the Most Promise for Improving Research Data Security?**

Option	Enterprise 2011 Responses	Dartmouth Responses
Expand data governance to explicitly include research data	5	6
Ensure regulatory compliance	0	1
Partner with the university research and compliance offices	4	0
Address research data management issues	3	4
Ensure research servers have competent systems administrators	5	4
Deploy enterprise/departmental firewalls	1	0
Establish a “DMZ” for research servers	1	0
Ensure researchers understand their responsibility in protecting data	6	4
Increased training for researchers*	n/a	3
<b>Total</b>	<b>25</b>	<b>22</b>

\* This question was added after the Enterprise 2011 presentation.

## What It Means to Higher Education

It's a cliché to say that the amount of data being generated in our personal and professional lives is exploding. One persistent descriptive phrase we like is “data tsunami,” as we think this captures the rapidity and overwhelming impact on our lives. Nowhere is this more evident than in the amount of research data produced at our institutions. Most disciplines today have at least some significant digital component, and many are fundamentally digital in their execution. Several of the traditional physical and social sciences have led the way in this regard, particularly genome sciences–related research and accompanying data needs. Although this bulletin focuses on research data, note that the issues also apply to broader areas of scholarship and artistic expression—for example, meeting the needs of the digital humanities and of digital performance capture.

Current grant funding and institutional library and IT budgets are simply not structured to address the longer term curation and preservation needs of our institutions, let alone the near-term storage requirements. Grants have a beginning and an end, typically with a finite and fixed amount of funding. However, funding entities, domain-based professional associations, and ongoing research and scholarship increasingly require longer term persistence, security, integrity, and availability of the digital resources that underlie scholarly activities beyond the end of a grant. Further, researchers and scholars are using existing data files in new and innovative ways that go beyond the purposes for which the data were originally created. What are the obligations of researchers to maintain and/or retain the data on which published articles and scholarship are based?

Addressing these issues may well infringe on the classic independence of the individual primary investigator and his or her team. Other entities must of necessity get involved, as PIs are simply not in a position to address these issues unilaterally or in isolation, as has historically been the case. In many ways, addressing these issues seems like an extension of the traditional library and central IT roles, so the need for a partnership between these two entities is clear. First, however, the cultural, funding, ownership, stewardship, and governance issues need to be sorted out. Clearly, the university's research office should be involved because these issues increasingly lie at the heart of research, scholarship, and funding, and providing whatever additional resources and services are necessary will need to become part of the institution's research-funding mechanisms and structures.

In addition—at least on our campus—it has come as a bit of shock to some researchers to find themselves accountable for the confidentiality, integrity, and availability of the data they generate and/or amass in the course of their scholarly endeavors. Some researchers have noted that they are not security experts, so how can they be held accountable for the security of their data? Our response is to point to the Internal Revenue Service: most of us are not tax attorneys or accountants, but the IRS still holds us accountable for submitting an accurate and legal tax return each year. It is our responsibility as taxpayers to get competent help. With this as a metaphor, researchers are similarly responsible to ensure they get competent guidance and support to ensure their data are protected. This can often be provided by the institutional IT organization. At UNC, we've utilized the Security Liaisons to communicate the kind of assistance ITS offers to researchers.

## Key Questions to Ask

- What are the policy issues associated with protecting the security of research data?
- What is the state of data governance at your institution? For example, who determines which data assets are to be protected, by whom, and for how long? Who should make those determinations?
- Who is responsible for which aspects of data security? What should researchers do? Funding agencies? The institution? IT?
- Securing information systems and resources is expensive. How can you fund this on a sustained basis?
- Do you know what your risks are in terms of confidentiality, availability, and integrity of data?
- What assurances do you have that your campus research systems storing sensitive and mission-critical data are adequately protected?
- What laws and compliance regulations govern your sensitive information?

## Where to Learn More

- Committee on the Human Dimensions of Social Change. Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data (2007). [http://books.nap.edu/openbook.php?record\\_id=11865&page=83](http://books.nap.edu/openbook.php?record_id=11865&page=83)
- Intel Science & Technology Center for Secure Computing. <http://blogs.intel.com/research/2011/06/istc-sc.php>
- Internet2 Guidelines for Data De-Identification or Anonymization. <https://wiki.internet2.edu/confluence/display/itsg2/Guidelines+for+Data+De-Identification+or+Anonymization>
- EDUCAUSE 2003. IT Security for Higher Education: A Legal Perspective. <http://net.educause.edu/ir/library/pdf/csd2746.pdf>
- Dept of Education 2011. Family Educational Rights and Privacy Act (FERPA). <http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>
- HHS 2011. Summary of the HIPAA Security Rule. <http://www.hhs.gov/ocr/privacy/hipaa/understanding/srsummary.html>
- FTC 2011. Gramm-Leach-Bliley Act. <http://business.ftc.gov/privacy-and-security/gramm-leach-bliley-act>
- NIST 2010. Guide for Applying the Risk Management Framework to Federal Information Systems. Special Publication 800-37. <http://csrc.nist.gov/publications/nistpubs/800-37-rev1/sp800-37-rev1-final.pdf>
- NIST 2009. Recommended Security Controls for Federal Information Systems and Organizations Special Publication 800-53. <http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final.pdf>

- NSF data management plan requirements.  
[http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg\\_2.jsp#dmp](http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp)
- UNC Research Data Toolkit.  
[http://www.lib.unc.edu/reference/data\\_services/researchdatatoolkit/index.html](http://www.lib.unc.edu/reference/data_services/researchdatatoolkit/index.html)
- UNC IT Security Policies. [http://its.unc.edu/ITS/events\\_and\\_projects/CCM3\\_020504](http://its.unc.edu/ITS/events_and_projects/CCM3_020504)

## About the Authors

*Larry D. Conrad is vice chancellor for IT and CIO at the University of North Carolina at Chapel Hill, where Ruth Marinshaw is assistant vice chancellor for research computing, and Stan Waddell is executive director and information security officer.*

### Citation for This Work

Conrad, Larry D., Ruth Marinshaw, and Stan Waddell. "Protecting the Security of Research Data." (Research Bulletin). Boulder, CO: EDUCAUSE Center for Applied Research, November 8, 2011, available from <http://www.educause.edu/ecar>.

### Copyright

Copyright 2011 EDUCAUSE and Larry D. Conrad, Ruth Marinshaw, and Stan Waddell. CC by-nc-nd

- 
1. National Science Foundation. Campus Bridging Task Force sponsored workshop, "Campus Leadership Engagement in Building a Coherent Campus Cyberinfrastructure." <http://pti.iu.edu/campusbridging/leadership>
  2. National Science Foundation. "Workshop on Research Data Lifecycle Management." <http://rcs.columbia.edu/rdlm>  
Ponemon Institute. "Fifth Annual US Cost of Data Breach". <http://www.ponemon.org/data-security>
  3. Ponemon Institute. "Fifth Annual US Cost of Data Breach". <http://www.ponemon.org/data-security>
  4. Joesph Campana. How safe are we in our schools. Retrieved from <http://www.jcampana.com/JCampanaDocuments/EducationSectorDataBreachStudy.pdf>
  5. Committee on Science, Engineering, and Public Policy (U.S.). Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. "Ensuring the integrity, accessibility, and stewardship of research data in the Digital Age." (Washington, DC: The National Academies Press, 2009), [http://www.nap.edu/catalog.php?record\\_id=12615](http://www.nap.edu/catalog.php?record_id=12615).
  6. Larry Conrad. "Out of the Frying Pan and into the Fire: Protecting the Security of Research Data." Presentation at the Educause Enterprise 2011 conference, Chicago, IL, May, 2011, available from <http://www.educause.edu/ent11>.
  7. Larry Conrad. "Out of the Frying Pan and into the Fire: Protecting the Security of Research Data." Presentation at Dartmouth Securing the e-Campus conference, Hanover, NH, July, 2011, available from <http://www.ists.dartmouth.edu/events/ecampus/>.