ECAR working groups are where EDUCAUSE members come together to create solutions to today’s problems and provide insight into higher education IT’s tomorrow. Individuals at EDUCAUSE member institutions are invited to collaborate on projects that address core technology challenges and advance emerging technologies important to colleges and universities. More information can be found at the ECAR working groups website.

Introduction

How does the modern IT organization provide the guidance, infrastructure, and support that are critical to innovation and entrepreneurial activities? How can IT strategies and policies support discovery and innovation while ensuring that data generated across academic, research, and business units at an institution are secured?

Data protection demands proactive consideration of numerous issues and cultural shifts, even as data protection tools and practices evolve. Contemporary challenges include personal data generated by wearable sensors; personal awareness, consent, anonymity, and data-sharing controls; data entry and storage; and the use of data in advanced algorithms and analytics. These examples represent only some of the countless types of data generated and handled on a daily basis at an institution of higher education today.

The variety and volume of institutional data are phenomenal, and data management is increasingly decentralized. Institutional data can include educational, research, medical, business, student, and employee data. These large data sets can easily become unstable and vulnerable to corruption, which can lead to a failure to keep the data accurate, complete, and whole.

What Is Data Protection?

This paper defines data protection as “the principles and processes for protecting data from corruption, misapplication, compromise, misuse, and loss.” Data protection leverages evolving technology and emerging practices to focus on how to apply the three critical information security imperatives: confidentiality, integrity, and availability.

These imperatives are not new, nor are they unique to data. Rather, they have applied to institutional records for as long as records have been kept, even printed on paper and stored in filing cabinets. As institutions move to paperless systems, they must view the same concerns through a digital lens, and they must ensure that the risks to business continuity are known and effectively managed.
Data protection can be organized into four domains (see table 1):

- **Affirmative Protections**: These help define the positive cultural stance an institution has toward data protection, and they indicate the investment that an institution has made in protecting data. They include governance, policy, and procedures that are developed proactively to support data protection.

- **Misuse Protections**: These focus on ensuring that privacy and security needs are met. Included here are protections that ensure compliance and regulatory requirements are met, as well as actions that need to be taken in case of a breach.

- **Business Continuity Protections**: These protections focus on the continued availability and usability of data.

- **Data Life-Cycle Protections**: These are the procedures and policies specific to all stages of the data life cycle.¹

Table 1. The four domains of data protection

<table>
<thead>
<tr>
<th>Affirmative Protections</th>
<th>Misuse Protections</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Governance</strong></td>
<td><strong>Privacy Rights</strong></td>
</tr>
<tr>
<td>Ensuring data conform to quality and standards</td>
<td>Ensuring personal awareness of rights and exercise of</td>
</tr>
<tr>
<td>upon entry and throughout use in operations and</td>
<td>personal choice in use of data</td>
</tr>
<tr>
<td>analysis; includes a decision framework around</td>
<td></td>
</tr>
<tr>
<td>what data should be collected and protected</td>
<td></td>
</tr>
<tr>
<td><strong>Integrity Assurance</strong></td>
<td><strong>Information Security</strong></td>
</tr>
<tr>
<td>Monitoring to ensure data are complete, accurate,</td>
<td>Application of controls to protect data and systems</td>
</tr>
<tr>
<td>and timely and conform to data standards</td>
<td>based on data classification; addressing storage media,</td>
</tr>
<tr>
<td></td>
<td>hardware, databases, operating systems, network and</td>
</tr>
<tr>
<td></td>
<td>transmission, and physical and cloud data centers</td>
</tr>
<tr>
<td><strong>Data Classification</strong></td>
<td><strong>Incident Response</strong></td>
</tr>
<tr>
<td>Determining and categorizing the sensitivity of</td>
<td>Procedures established and tested for responding to a</td>
</tr>
<tr>
<td>data and necessary protections</td>
<td>breach in a timely manner, consistent with policy</td>
</tr>
<tr>
<td><strong>Access Management</strong></td>
<td><strong>Data-Sharing Agreements</strong></td>
</tr>
<tr>
<td>User identification, authorization, and</td>
<td>Documented agreements for exchange and use of data,</td>
</tr>
<tr>
<td>authentication to access systems and data</td>
<td>internally between units and with external vendors and</td>
</tr>
<tr>
<td></td>
<td>others</td>
</tr>
<tr>
<td><strong>Culture of Data Value</strong></td>
<td></td>
</tr>
<tr>
<td>Transforming institutional mindset to value data</td>
<td></td>
</tr>
<tr>
<td>as an asset, under the direction of data stewards</td>
<td></td>
</tr>
<tr>
<td><strong>Business Continuity Protections</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Backup</strong></td>
<td><strong>Collection</strong></td>
</tr>
<tr>
<td>Routine, scheduled backups of critical data to</td>
<td>The process of bringing data into the life cycle in a</td>
</tr>
<tr>
<td>alternate hardware or cloud; may entail off-site</td>
<td>deliberate and thoughtful way</td>
</tr>
<tr>
<td>storage by a third party</td>
<td></td>
</tr>
<tr>
<td><strong>Change Management</strong></td>
<td>**Thorough, scheduled stakeholder notification and</td>
</tr>
<tr>
<td></td>
<td>prior acceptance testing of changes to databases and</td>
</tr>
<tr>
<td></td>
<td>systems to avoid data corruption and loss</td>
</tr>
<tr>
<td><strong>Systems Redundancy</strong></td>
<td><strong>Archive</strong></td>
</tr>
<tr>
<td>Technical architecture, including virtualization,</td>
<td>Removal of records from production systems and retention</td>
</tr>
<tr>
<td>to ensure critical systems remain available</td>
<td>for a specified time period for access as needed</td>
</tr>
<tr>
<td><strong>Disaster Recovery</strong></td>
<td><strong>Deletion</strong></td>
</tr>
<tr>
<td>Procedures established and periodically tested for</td>
<td>Permanent removal of records from systems, hardware,</td>
</tr>
<tr>
<td>restoring systems from backup in the event of a</td>
<td>and media</td>
</tr>
<tr>
<td>natural or manmade disaster</td>
<td></td>
</tr>
<tr>
<td><strong>Disposal</strong></td>
<td>**Permanent, unrecoverable destruction of records and</td>
</tr>
<tr>
<td></td>
<td>storage media/hardware</td>
</tr>
</tbody>
</table>
The digital era has, however, amplified a debate over who “owns” data; because electronic data can be more readily shared and accessed than paper records, the answer to the question of who has responsibility for data is less clear. Moreover, although an institution is the owner of most data it archives, institutions may also store data owned or controlled by others who require unique data protections. For example, government agencies and foundations awarding grants for sponsored faculty research may require specific data protection controls or seemingly contradictory requirements to make data publicly available. In the United States, federal and industry compliance controls include:

- Federal Education Rights and Privacy Act (FERPA)—student records
- Gramm-Leach-Bliley Act (GLBA)—student loans
- Health Information Portability and Accountability Act (HIPAA)—protected health information
- Defense Federal Acquisition Regulation Supplement (DFARS), International Traffic in Arms Regulations (ITAR), and Export Administration Regulations (EAR)
- Payment Card Industry Data Security Standard (PCI DSS)—industry standard

Campus users demand fast access to data. They expect data to be readily comprehensible, usable, and of high quality and integrity. At the same time, users must apply stronger privacy standards and security controls. This is the complicated campus ecosystem that is “data protection.”

**Major Shifts**

The constant evolution of technology challenges higher education institutions to continuously refine data protections. Broad environmental and cultural shifts in user expectations and innovative uses for data present additional challenges for appropriate data protection that balances demands for confidentiality, integrity, and availability of data, now and into the future.

**Environmental Shifts**

Four major environmental shifts have introduced confusing, complex, and seemingly contradictory challenges for data protection. One such shift is the proliferation of diverse devices, including computers, handhelds, and even wearable, networked, data-producing technology; these devices have produced an almost immeasurable increase in the volume of data. The availability of these data has in turn produced new and legitimate uses for data under the umbrella of business intelligence and analytics, as well as theft and data breaches.

**Device Proliferation**

Data are no longer solely the product of computers that an institution owns, maintains, and protects. While those assets still exist, students, faculty, staff, and visitors now bring multiple devices of their own to campus, including personal computers, tablets, mobile phones, gaming systems, wearable sensors, and medical devices. Further, Internet-connected appliances including environmental and infrastructure controls have now become routine members of network and computing infrastructures. New terms—Internet of Things (IoT) and Industrial Internet of Things (IIoT)—have emerged to describe these devices. In fact, the concept is so broad that it is now being termed as the “Internet of Everything” (IoE). Users expect to connect their personal IoEs seamlessly with campus resources, networks, and data, even as their IoEs rapidly produce more data, much of which has nothing to do with the normal business concerns of a college or university.
Proliferation and Location of Data

“Rapid” barely describes the production of data in recent years. The volume of data alone makes on-campus storage costly and impractical. The resultant shift to cloud storage and computing impacts the level of control institutions are able to provide for institutional data. We now frequently rely on cloud service partners to protect institutional data on our behalf. To delegate that responsibility requires new tools and practices: third-party vendor assessments; contractual arrangements; data sharing, protection, business associate agreements; and breach notification clauses.

However, the infrastructure, services, and controls that cloud services provide cannot fully ensure the confidentiality of the data. Institutions are responsible for maintaining access control authorization. On the one hand, sharing of information is a significant strength of cloud storage services; however, that capability raises data protection concerns: Do we know what sensitive information is being shared, and do current access-revocation processes appropriately address data stored in the cloud?2

Though the volume of data is increasing and may be spread across various locations, stakeholders still expect their data to be safe, secure, and easily accessible on demand. Historically, data backups have been used to meet this expectation. But restoration from backup could take weeks in the event of a large-scale disaster. The newer approach is to store copies of data in the original format off-site, potentially in the cloud. As cloud services mature, service level agreements (SLAs) between the university and the cloud vendor must ensure appropriate services for backup and recovery.

New Data Uses

In keeping with their historic role of innovation, colleges and universities continue to envision new and creative uses for data. Efforts to modernize campus services; to attract and recruit potential students, employees, and donors; and to produce optimal educational outcomes and research can all benefit from the appropriate use of institutional data. “Big data” is the term commonly used to describe the deluge of data in today’s networked, digitized, sensor-laden, and information-driven world, and “analytics” is the contemporary term for data analysis, insight, and business intelligence. Big data is often characterized by “the four Vs”3 and requires a scalable architecture and specialized tools for efficient storage, manipulation, and analysis:4

- **Variety**: different forms and subjects of data
- **Volume**: large, measurable quantity and scale
- **Velocity**: rapid production and streaming delivery
- **Veracity**: certainty about the accuracy of data and its subject

New Threats, New Exposure

Data protection threats are not new to higher education, but the threat landscape is always changing. One theory is that as banking, defense, and retail sectors invest in data and can constrain what devices or software can be used on their networks, a resource-constrained industry such as education may present a softer target because it allows students and researches to use a broad range of technology and
collaborate globally. Current threat actors consist of hacktivists, hackers, cybercriminals, and nation-state actors. Their actions have become more noticeable and publicized. Attack vectors and targets now frequently include distributed denial of service (DDoS) attacks, theft of personally identifiable information and protected health information, and attacks by nation-state actors. Institutions must undertake efforts to reduce their level of risk, lest they become the easiest targets.

For many institutions, facilitating data protection requires a shift in thinking and new investment in infrastructure, processes, and resources. Defense in depth, the concept of utilizing layered security controls for data protection, is one framework that has been adopted for use in higher education. We must now recognize that defense in depth, on its own, is not good enough. Dedicated threats such as organized cybercriminal gangs and nation-state actors will get into our systems. We must be able to recognize when we have been breached and remove the threat before significant damage is done by interrupting the cyberattack chain.

**Cultural Shifts**

Data entry by way of personal computers began transforming campus operations in the 1980s. Data—and responsibility for them—became democratic. Data supports and derives from the many administrative and business functions of a university: admitting, enrolling, teaching, and housing students, as well as hiring and paying employees, and using online self-service functions. Data can overwhelm faculty researchers, who collect, produce, report, and exchange data in their teaching, service, and research roles. Protection of data has become equally democratic, though perhaps without the degree of focus and investment truly needed. It is an institution’s responsibility to ensure that administrators, researchers, and faculty alike follow procedures and protocols.

The cultural transformation around data extends to ever-increasing expectations that organizational units not only understand the value of the data they create, use, and oversee for their own purposes but also appreciate that their data may be a valuable asset to other departments and the institution.

**The Value of Data**

As an asset, data require protection in the same sense that physical property and other assets need protection. Both require practices and processes that ensure institutional access while preventing theft and misuse. Contemporary culture recognizes that data are intrinsically valuable to the institution and that data loss and misuse are consequential. Contemporary culture is also recognizing that data can be monetized. The monetary value of administrative data exists for organizational units and for the institution as a whole. Faculty-produced research data may also be monetized and their value augmented when “mashed up” with other data sources.
Protection is an important element of preserving the value of data assets. Further acknowledgement of the need for a data protection emphasis came in the 2016 ECAR Top 10 IT Issues list, which identified “Improving the management of institutional data through data standards, integration, protection, and governance” as a top issue. In addition to security, the proliferation of privacy concerns represents another cultural change.

Including Privacy with Protection

In recent years, individuals, institutions, national governments, and multinational organizations have increasingly emphasized the principles of personal privacy. Privacy considers what data are collected, from whom, for what purposes, where they will be stored, and how they will be used. Moreover, privacy principles have established the right of individuals to have their records removed from data sets or withheld from certain uses.

In October 2015, the European Court of Justice found that the “right to be forgotten” is a fundamental human right, but the ruling applies only to search engines. As a result, individual Europeans can apply to be delisted from Google’s search index. Google is not currently required to delist such search results globally—only for results in the European country where the request originated. While this doesn’t apply currently in the United States, constantly changing legal decisions both in the United States and around the world require that institutional practices and policies be continuously reviewed and adjusted accordingly. Nuances such as this not only illustrate the complexity of deciphering and implementing data protection efforts but also highlight the cultural and legal importance data protection is being given. Data protection poses challenges, but not necessarily barriers, to the use of data for business intelligence and competitive advantage.

A New Data Economy

The emergence of cloud infrastructures and associated access enhancements have dramatically lowered private-sector costs for storing data and making them available for widespread use. Private companies now provide services that generate and store data previously produced by and housed in institutions, but their business models are beyond the influence of an individual institution. These models operate efficiently by providing similar services across many institutions, potentially combining data for analysis and business intelligence that they in turn sell to the same institutions.

Companies such as D2L (learning management systems), EAB (student success advising), Udemy (online course development), and Knewton (adaptive learning systems) are creating actionable knowledge from each new data product they introduce. Institutional data protection initiatives are under pressure to adapt to, and co-exist with, such business models as universities opt to buy rather than build these infrastructures and services. A recent article in *EDUCAUSE Review* contends that one of the likely roles of the future IT organization will be to generate and host massive intra- and interinstitutional research data sets that may also include joint research with private entities. It is likely that these collaborative data sets will not only be valuable to the researchers involved but—in the case of new product development—might also have significant monetary impact to the entire institution.
Conclusion

As a result of today’s dynamic environment, the definition of “data protection” is changing. What does it mean to protect data now and into the future, not only when data are in transit but also when untold volumes of data aren’t even stored locally? How do we determine what data to protect and when to no longer protect (and/or destroy) data, and how do we identify what business needs are being met?

Environmental considerations, culture, and data protection practices are constantly changing and evolving, and so too must the data protection initiatives employed by IT and data stewards. The traditional approaches to data protection no longer scale and do not meet current expectations. Backing up to tape or multiple disks no longer addresses the wide variety of data types and security requirements that exist, nor does it recognize factors such as geographic diversity and the effects of cloud services in this space. The proliferation of data to externally provided services diminishes our capability to ensure that data are protected with institutional resources and according to law and policy. An expanded focus on contractual protections is needed as we continue to migrate services to the cloud and rely on these external services.

Data protection is a broad subject area that requires the input of stakeholders such as information security, library, and IT organizations, among others, and needs to be endorsed and supported from the top down. Based on the perspectives of IT, senior leadership, and business, there is a need for a broad conversation to understand requirements within this space and to determine how we can work together to adjust institutional culture and put into place practices and services to protect the data and the institution.

Authors

Special thanks go to the following ECAR Data Protection (ECAR-DP) Working Group members who authored this report:

Michael Fary
Enterprise Data Architect
University of Chicago

Michael Flanigan
Director of Planning and Operational Technologies
Virginia Commonwealth University

Dale Joseph
Manager, CTS Storage Systems, Operations, and Recovery
Duquesne University

Karuna Joshi
Research Assistant Professor, Computer Science and Electrical Engineering
University of Maryland, Baltimore County

Michael C. Kelly
Chief Data Officer
University of South Carolina

Mike Lowry
Assistant Director of Information and Infrastructure Assurance Operations
University of Michigan

Aprille C. McKay
Assistant Director of University Collections and Records Management
University of Michigan—Ann Arbor

Geoffrey S. Nathan
Policy and Privacy Coordinator
Wayne State University

Kim Owen
Advanced Applications Outreach
North Dakota State University

Theresa Semmens (ECAR-DP Chair)
Chief Information Security Officer
North Dakota State University
In addition, the following ECAR-DP Working Group members contributed to the report:

**Jason Fishbain**  
Chief Data Officer  
University of Wisconsin–Madison

**Gary Starling**  
Chief Information Security Officer  
University of Denver

**Keith Schoenefeld**  
Information Security Director  
Elon University

**Clare van den Blink**  
Vice President and CIO  
Pace University

**Jamie Spradlin**  
Chief Information Security Officer  
University of North Carolina at Pembroke

---

**Citation for This Work**


**Notes**

1. For more on the data life cycle, see Michael Fary and Kim Owen, *Developing an Institutional Research Data Management Plan Service*, EDUCAUSE, January 2013. Figure 1 in this document describes common data life-cycle stages, from data creation to data reuse and transformation.

2. More information about cloud services can be found in the following ECAR working group reports: *Preparing Your IT Organization for the Cloud (Series)*, *Research Computing in the Cloud*, and *TCO for Cloud Services: A Framework*.

3. See “The Four V’s of Big Data,” IBM, infographic. “Big data” has been defined in many ways. In 2013, in the *ACTI Data Management Glossary*, the EDUCAUSE Data Management Working Group defined it as “a loosely defined term used to describe data sets so large and complex that they become difficult to process and store using data management tools.” In 2014, the ECAR Campus Cyberinfrastructure Working Group launched a series focusing on *Big Data in the Campus Landscape*; in the introduction, they looked at the defining characteristics of big data, noting that “perhaps the one consistency across these definitions is simply that big data is data that is ‘big’ and doesn’t readily lend itself to conventional IT practices or casual treatment.”


