In his role as special-projects coordinator, Joe had seen assignments of all shapes and sizes. So when his boss Claire, vice president for finance and operations, said, “This is a big one,” he listened closely.

“Joe, the last audit turned up two significant findings. They’re both related to data governance. I don’t think these are isolated issues. Our data governance and stewardship policies and processes haven’t been seriously reviewed in years. I think this is a good time to refresh these, and I want you to lead this effort. I’ll send you the audit findings, and some people to talk to. Look at the existing policies and processes, have some conversations with stakeholders, and come back to me with a rough proposal. Two weeks?”

Headlines reporting costly, high-profile security breaches and data losses are all too common, and higher education is not immune to such threats. In fact, with traditions of open access, knowledge sharing, and
decentralized governance, higher education institutions can be more vulnerable than other types of organizations to data-related risks. All the hardware and firewall rules in the world are not going to protect an institution’s data if the institution does not know what data exist, where the data are located, and who is responsible for the safekeeping of the data. Good data management practices must be followed. This paper is intended to serve as a conversation starter among stakeholders and executives about the importance of having well-planned data governance. It identifies the risks of not having effective governance and the benefits of a good, well-implemented governance strategy, and it outlines the essential purposes and activities of a data governance program, answering three questions: What is data governance? Why is data governance necessary? Who should be involved, and how?

What Is Data Governance?

Data governance assists in maintaining data integrity, controlling access, and securing data storage for the institutional asset. A data governance program seeks to identify what data currently exist or need to exist in the future, and it helps identify disparate data silos on campus. Such a program defines roles and responsibility for managing data and assigns accountability to specific groups or individuals within the institution through a strong data stewardship model. Although an enterprise perspective should govern an institution’s data, an effective governance plan does not necessarily centralize control of the institution’s data; rather, it creates systems that identify institutional and departmental data managers and provides guidelines for these managers as they handle data.

By creating these systems, a data governance program mitigates important risks associated with the storage and management of data. The costs associated with accidental disclosure of private data are well known. But how are “private data” defined, and how certain can an institution be that all of the people with access to data understand this definition and are prepared to apply it? Further, how can the institution be certain that data are not altered by untrained or unauthorized parties in ways that compromise the quality of the data—and with it the quality of the decisions made using the data? A data governance program reduces these risks by creating systems and processes under which data are only handled by individuals who have been properly trained to handle them.

Why Is Data Governance Necessary?

Establishing and maintaining good data governance will require time, talent, and funding, all of which demand solid justification: Why is data governance important? The answers fall into two categories: benefits (positive deliverables) and risk management (reduction of negative consequences). Functional
offices such as the registrar and human resources often have more control over their own domains and
stand to derive greater value from both benefits and risk management, making partnerships essential to
data management.

Later, Joe reviewed the audit findings related to the deployment of a new student-housing system. This
application originally had two sets of users: student affairs staff (full access) and students (limited access).
Finance personnel requested the ability to look at raw student transactional data, which were more detailed in
the housing system than in the central finance system. Developers responded to this request by creating a
new type of “finance” student user with access only to student financials. This configuration opened a path for
student users of the system to access other students’ financial information under some conditions. No
inappropriate access was discovered, and the vulnerability was corrected (at the cost of finance losing access
to the more detailed transactional data).

That afternoon, Joe talked to his colleague Jen in student affairs about how they use institutional data. “Well,”
she said, “we’ve gotten pretty good at identifying academically at-risk students. We’re now beginning to run
satisfaction surveys with students and correlate responses to their engagement with specific departments,
programs, instructors, extracurricular activities, etc. Once we identify outliers, we plan to follow up with more
detailed inquiries. We’re open to surprises—our early findings, for example, suggest some satisfaction is
based on housing. Some dormitories score higher than others, and now we are looking at the results to see if
they break down differently by gender, country of origin, or ethnic background, for instance. We want our
students not just to do well academically but also to enjoy their time here. To do that we need to know what’s
working for them, and what’s not.”

Benefits are positive outcomes or deliverables, including but not limited to:

- **Official versus ad hoc data definitions.** For official reporting purposes, a data element such as a
  person’s race/ethnicity should be defined consistently; the definition needs to be used across
  campus; and the system to contain the authoritative record of that definition needs to be determined.
  In other situations, alternative definitions of a data element need to be used for ad hoc analysis. A
data governance program should establish the mechanisms by which the official definition is derived,
communicated, and accessed by all constituents, as well as the mechanisms for allowing differing but
necessary definitions to exist in other types of analyses. The governance program should ensure that
all reports that should contain the official definition do so and that those analyses that do not use the
official definition clearly state that fact.

- **Clear responsibilities.** For specific data elements—for instance, a Social Security number—data
governance clarifies which departments are responsible for initial entry and maintenance of data (the
data steward), how sensitive the data element is (security classification), who has access to the data,
how changes to data definitions will be communicated, and under what circumstances the data
element may be used—a practice generally referred to as data stewardship.

- **Structured and unstructured data.** As technology continues to evolve, the capabilities to analyze
different types of data (i.e., structured versus unstructured data) will mature. A data governance
program should recognize this trend and establish the appropriate processes, best practices, and
tools to enable the potential of data from all types. Data governance should determine and clearly
communicate which data elements must be captured and consistently assigned in a standardized,
structured format and, conversely, when leaving data in an unstructured format is the desired
approach. In either case, policies around how each type of data is handled should be created and
enforced through data governance.
Musing over the risk aspects of the issue, Joe met his friend Lou for lunch. Lou worked in the university’s research administration office. As Joe started to explain his new project, Lou smiled. “We face many of the same issues,” he said. He talked about increasingly strict federal regulations around data sharing, the complexities of working with the wide variety of data types created by researchers in various fields, and the cost and relative immaturity of holistic solutions to research data management problems. However, as they talked, Joe noticed that although the problems were the same in the abstract, they were very different in their implementation—the regulatory agencies, the specifics of their regulations, the technical solutions, and the individuals and organizations that needed to collaborate across campus all differed for research data and administrative data.
Research data management shares many concepts with data management in the administrative domain, but the specifics are quite different. For most organizations, these problems will be tackled by different groups of people; ideally, organizations will be able to foster sharing of experiences and best practices between these efforts.

Though some governance issues are common to administrative and research data, it is important to recognize the many differences in how data are obtained, stored, retrieved, and protected. These variations need to be considered when outlining a governance roadmap. Administrative data (for example, from HR, financial, employee, or student systems) are typically generated from transactional systems; as such, the rules for protection and preservation are relatively consistent. Research data sets, on the other hand, may consist of many different sources, aspects of which may be very sensitive; this requires more training and oversight of people and processes to ensure they protect and store data appropriately. Furthermore, the compliance regulations for grant-awarding processes are growing in complexity and vastly differ depending on the type of data being used and how they are collected.

**Who Should Be Involved, and How?**

Who should be involved in data governance at any institution is a complicated matter, requiring awareness of internal politics, priorities, and personalities. Recognizing three constituent groups, and involving them in the right way, at the right moments, is key.

The first constituents are **operational area managers**, who live with data on a daily basis. They recognize pain points with data, have a keen understanding of business processes and data life cycles, and can envision and implement improvements. These operational managers are natural candidates to form an initial team to discuss and advocate for data governance. This team might draft a proposal addressing what should be done differently; outline the objectives, direct benefits, and risks that will be managed; and specify resources needed.

The second constituents are **institutional executives**. Leaders are accustomed to receiving data in reports, such as dashboards and analytics; however, leaders typically do not work with the records and data directly. As a result, they may not be fully aware of the challenges and level of effort necessary to produce the reports they receive. Educating university leaders on the need for clear data governance structures is thus a key step in developing and implementing a data governance strategy. Leaders should be presented with data governance issues in terms that make sense to them—

---

**Cost Considerations**

Data governance cannot be achieved without resources, some of which are not self-evident. Two areas of potential cost are:

- **Personnel.** Staff from core operational and administrative areas with daily responsibility for data will need to be actively engaged in data governance. Some new personnel may be needed; some existing staff may need their responsibilities and time allocation adjusted.

- **Data governance tools.** Off-the-shelf software and information systems are available to facilitate data governance activities and services, such as data glossary, workflows for data access requests, master data and reference data management, inventory of information systems and available data sets, and data-quality monitoring. Such a repository expedites access, clarifies responsibilities, and enables new initiatives like analytics and data visualization.

Securing commitments and funding from institutional executives for these resources may be key to the data governance program’s long-term success and sustainability.
how have previous issues complicated reporting, analysis, and compliance with deadlines? A strong value proposition is key; it must be clear how commitments of current or new resources will demonstrably improve outcomes or make new deliverables possible. A strong data governance strategy can only be implemented with executive endorsement, backed by policy and firm resource commitments; it is therefore critical to obtain such endorsement before taking concrete actions.

The third and final constituent group is the university community. The culture of higher education tends to require participatory decision making across multiple levels—faculty, staff, students, and administrators—and data governance benefits greatly from this inclusivity. All stakeholders who have a role in creating, maintaining, analyzing, or reporting data have a vested interest and must be engaged in data governance. These stakeholders should be directly involved in the development of the data governance strategy, especially in designing feedback loops for end users to raise data concerns or suggest changes to business processes. Further, the community should help identify the right people for key roles; prioritize the activities and services of a data governance program; and develop a timeline for implementing and maturing data governance.

A few days later, after conversations with other campus partners and a review of the existing policies and procedures, Joe stopped by Claire’s office. “We have a lot of work to do,” he said, “but I think we have a pretty clear idea of what the path looks like.” Claire smiled. “Thanks.”

Conclusion

Governance is not a new concept to higher education. To some extent, data governance has been in place, more or less de facto, for years. Student data, personnel data, faculty data—all have been claimed by different areas of the institution for their care and use.

As institutions understand the immense value in the data—as an asset as well as a tool—governance becomes increasingly important. As with any asset, management of the data, the processes that create and manipulate it, and the ability to use it for added value are key to the governance process.

Along with the perceived value of the data, increasing regulation and risk management oversight are being implemented to protect the institution and its faculty, staff, and students. Research data and intellectual property are important revenue generators that institutions fiercely guard. In such cases, an institution’s data governance practices can mean the difference between maximizing the value contained in the data and allowing a valuable asset to be wasted or, even worse, misused.

In addition, it is important to ensure the ongoing education of the faculty, staff, and other constituents about data governance processes and imperatives. At the same time, instituting a continued review of new policies, laws, and practices that have an impact on data governance is essential.

Data governance is a shared university imperative. Understanding the what, why, and who is the first step in establishing a new or enhancing an existing data governance program. To effectively manage their data, many institutions are establishing a variety of data management roles in the governance process. A new working group from the EDUCAUSE Center for Analysis and Research is being formed to address this next step by providing guidance on defining a data stewardship program, including a discussion of
the different types of data stewards and managers, how they work with colleagues to ensure that data are maintained, and necessary special skills or training.¹

Authors and Contributors
Special thanks go to the following ECAR working group members.

Authors

Douglas Blair
IT Consulting Engineer
Carnegie Mellon University

Kelly Briner
Director of Data Management
Arizona State University

Vaibhav Dani
Assistant Director, Database Architecture
Old Dominion University

Michael Fary (Co-Chair)
Enterprise Data Architect
University of Chicago

Jason Fishbain
Chief Data Officer
University of Wisconsin–Madison

Mylite D. Hart
Senior Systems Analyst
Rush University

Brian W. Hopkins (Co-Chair)
Director, Mississippi Center for Supercomputing Research
University of Mississippi

Michael C. Kelly
Chief Data Officer
University of South Carolina

Kathryn Matuch
Associate Vice President, Core Enterprise Systems
Drexel University

Eduardo Zaborowski
Director of Research Computing, Albert Einstein College of Medicine
Yeshiva University

Contributors

Ellen Amsel
Chief Information Security Officer
Princeton University

Nickolas Backscheider
Associate Executive Director, OIT
Auburn University

Gregory Orcutt
Senior Blackboard Administrator
The University of Texas Health Science Center at Houston

Note
1. More information about this ECAR working group project, “Data Owners: Establishing Stewardship Models,” is available online.